

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2001, 11:45:45 ; Search time 16.02 seconds
(without alignments)
1521.783 Million cell updates/sec

Title: US-09-609-383-2

Perfect score: 6317
Sequence: 1 MWGTRKAWVSFLVLEVTSL.....QSGVSLRPRVQAQPLIN 1184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 197339 segs, 20590346 residues

al number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6297	99.7	1184	2 US-08-918-914-1	Sequence 1, Appli
2	6297	99.7	1184	3 US-08-996-083-3	Sequence 3, Appli
3	3091.5	48.9	1156	3 US-08-996-083-1	Sequence 1, Appli
4	3091.5	48.9	1156	4 US-09-429-516-1	Sequence 1, Appli
5	3091.5	48.9	1156	4 US-09-429-516-3	Sequence 3, Appli
6	2898.5	45.9	599	2 US-08-954-333-7	Sequence 7, Appli
7	157	2.5	1170	1 US-08-313-288B-20	Sequence 20, Appli
8	147.5	2.3	788	2 US-08-918-914-4	Sequence 4, Appli
9	144	2.3	1911	1 US-08-348-006B-5	Sequence 5, Appli
10	144	2.3	1911	2 US-08-800-825A-5	Sequence 5, Appli
11	144	2.3	1911	4 US-09-158-657-5	Sequence 5, Appli
12	144	2.3	1911	5 PCT-US94-10166-5	Sequence 5, Appli
13	143	2.3	1172	1 US-08-313-288B-19	Sequence 19, Appli
14	142.5	2.3	4544	1 US-08-469-486-52	Sequence 52, Appli
15	142.5	2.3	4544	2 US-08-469-658-52	Sequence 52, Appli
16	138	2.2	1101	2 US-08-469-537A-96	Sequence 96, Appli
17	136	2.2	1118	1 US-07-934-393B-2	Sequence 2, Appli
18	136	2.2	1118	1 US-08-278-089A-2	Sequence 2, Appli
19	136	2.2	1118	2 US-08-838-957A-2	Sequence 2, Appli
20	135.5	2.1	898	2 US-08-808-982-5	Sequence 5, Appli
21	135	2.1	1122	2 US-08-278-089A-6	Sequence 6, Appli
22	135	2.1	1122	2 US-08-838-957A-6	Sequence 6, Appli
23	132	2.1	218	3 US-08-985-526-1	Sequence 1, Appli
24	132	2.1	239	5 PCT-US93-01652-1	Sequence 1, Appli
25	130.5	2.1	644	5 PCT-US93-00031-21	Sequence 21, Appli
26	130.5	2.1	1101	3 US-08-986-485-2	Sequence 2, Appli
27	128.5	2.0	807	1 US-07-862-021B-10	Sequence 10, Appli

28	128.5	2.0	807	1 US-08-313-288B-10	Sequence 10, Appli
29	128.5	2.0	807	5 PCT-US93-03164-10	Sequence 10, Appli
30	128.5	2.0	1338	3 US-08-750-141A-3	Sequence 3, Appli
31	128	2.0	1501	2 US-08-447-464-3	Sequence 3, Appli
32	128	2.0	1501	2 US-08-716-679-3	Sequence 3, Appli
33	127	2.0	441	3 US-08-985-526-3	Sequence 3, Appli
34	126	2.0	736	5 PCT-US93-00031-15	Sequence 15, Appli
35	125	2.0	643	5 PCT-US93-00031-19	Sequence 19, Appli
36	124.5	2.0	1018	1 US-08-408-093-6	Sequence 6, Appli
37	124.5	2.0	1018	1 US-08-408-420A-6	Sequence 6, Appli
38	124.5	2.0	1018	1 US-08-714-901-6	Sequence 6, Appli
39	124.5	2.0	1018	3 US-08-040-741-6	Sequence 6, Appli
40	121	1.9	596	2 US-08-752-307B-13	Sequence 13, Appli
41	121	1.9	802	1 US-07-862-021B-12	Sequence 12, Appli
42	121	1.9	802	1 US-08-313-288B-12	Sequence 12, Appli
43	121	1.9	802	5 PCT-US93-03164-12	Sequence 12, Appli
44	120.5	1.9	735	5 PCT-US93-00031-13	Sequence 13, Appli
45	120	1.9	392	2 US-08-799-173A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-918-914-1
Sequence 1, Application US/08918914
Patent No. 5876963
GENERAL INFORMATION:
APPLICANT: Mitchell, Peter
APPLICANT: Hutchinson, Nancy
APPLICANT: Lawton, Michael
APPLICANT: Magna, Holly
APPLICANT: Yocum, Sue
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/918,914
FILING DATE: Filed herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0369
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1184 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: ???
CLONE: 422069
US-08-918-914-1

Query Match 99.7%; Score 6297; DB 2; Length 1184;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1181; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGTGKAVSFLEVTSLVGRQTMLTQSVRRVOPGKKNPSIFAKPADTLESFGEWTTLF 60
 1 MGTGKAVSFLEVTSLVGRQTMLTQSVRRVOPGKKNPSIFAKPADTLESFGEWTTLF 60
 Db 1 MGTGKAVSFLEVTSLVGRQTMLTQSVRRVOPGKKNPSIFAKPADTLESFGEWTTLF 60

QY 61 NIDYGGKGYERLDAIRFYGGDRCARPLRLRARTDWTDPAGSTGVVHSGPREGFWCL 120
 61 NIDYGGKGYERLDAIRFYGGDRCARPLRLRARTDWTDPAGSTGVVHSGPREGFWCL 120
 Db 61 NIDYGGKGYERLDAIRFYGGDRCARPLRLRARTDWTDPAGSTGVVHSGPREGFWCL 120

QY 121 NRORPQONCSNTVRLCPGSLRRDTERIWSBPSWMSKSAACGQTVQTRTRICLAE 180
 121 NRORPQONCSNTVRLCPGSLRRDTERIWSBPSWMSKSAACGQTVQTRTRICLAE 180
 Db 121 NRORPQONCSNTVRLCPGSLRRDTERIWSBPSWMSKSAACGQTVQTRTRICLAE 180

QY 181 MVSICSEASEGQHCKQDCTACDLTCPMGOVNADCDACACODPMLHGAVSLPGAPASG 240
 181 MVSICSEASEGQHCKQDCTACDLTCPMGOVNADCDACACODPMLHGAVSLPGAPASG 240
 Db 181 MVSICSEASEGQHCKQDCTACDLTCPMGOVNADCDACACODPMLHGAVSLPGAPASG 240

QY 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKYKFAPIYLTMTKTSIKAATIK 300
 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKYKFAPIYLTMTKTSIKAATIK 300
 Db 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKYKFAPIYLTMTKTSIKAATIK 300

QY 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCATGKPRPDXYFWYHNDTLDPISLYKHESK 360
 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCATGKPRPDXYFWYHNDTLDPISLYKHESK 360
 Db 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCATGKPRPDXYFWYHNDTLDPISLYKHESK 360

QY 361 LVLRKLDQOHAGEFYCKAQSADAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDFQON 420
 361 LVLRKLDQOHAGEFYCKAQSADAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDFQON 420
 Db 361 LVLRKLDQOHAGEFYCKAQSADAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDFQON 420

QY 421 ATNSFYVDVRCPEVTKCAGQDNGIRCRDAVONCCISKTEEREIQCSTYLLPTKVAKEC 480
 421 ATNSFYVDVRCPEVTKCAGQDNGIRCRDAVONCCISKTEEREIQCSTYLLPTKVAKEC 480
 Db 421 ATNSFYVDVRCPEVTKCAGQDNGIRCRDAVONCCISKTEEREIQCSTYLLPTKVAKEC 480

QY 481 SCORCTETRSIVRGVSAADNGEPMRFGHAYMGNSRVSMTGKOTFTLHWPDTERLVL 540
 481 SCORCTETRSIVRGVSAADNGEPMRFGHAYMGNSRVSMTGKOTFTLHWPDTERLVL 540
 Db 481 SCORCTETRSIVRGVSAADNGEPMRFGHAYMGNSRVSMTGKOTFTLHWPDTERLVL 540

QY 541 FVDRLOKFEVNTTVLPNNKGSVNFHEIKMLRKEPITLAMEFNITPLGEVNGEDPMAE 600
 541 FVDRLOKFEVNTTVLPNNKGSVNFHEIKMLRKEPITLAMEFNITPLGEVNGEDPMAE 600
 Db 541 FVDRLOKFEVNTTVLPNNKGSVNFHEIKMLRKEPITLAMEFNITPLGEVNGEDPMAE 600

QY 601 LEIPSRSFYQNGEPTVIGKVKASYTFELDPNNISTATAQTDLNFINDEGDTFPLRTYGMF 660
 601 LEIPSRSFYQNGEPTVIGKVKASYTFELDPNNISTATAQTDLNFINDEGDTFPLRTYGMF 660
 Db 601 LEIPSRSFYQNGEPTVIGKVKASYTFELDPNNISTATAQTDLNFINDEGDTFPLRTYGMF 660

QY 661 SVDFRDEVTSEPLNAGVKYVHLSTOVKMPBHISTVYKLSINPDTGLMEBEGDFEFENOR 720
 661 SVDFRDEVTSEPLNAGVKYVHLSTOVKMPBHISTVYKLSINPDTGLMEBEGDFEFENOR 720
 Db 661 SVDFRDEVTSEPLNAGVKYVHLSTOVKMPBHISTVYKLSINPDTGLMEBEGDFEFENOR 720

QY 721 RNRREDTFLVGNLEIRERLFLNDVPESRCFVKVAYSERFLPSEOLOGVIVISINL 780
 721 RNRREDTFLVGNLEIRERLFLNDVPESRCFVKVAYSERFLPSEOLOGVIVISINL 780
 Db 721 RNRREDTFLVGNLEIRERLFLNDVPESRCFVKVAYSERFLPSEOLOGVIVISINL 780

QY 781 EPRTGLSNBRAMGRFDSVTGPNGACVPAFCDDQSPDAYSAVYLAISLAGEBLOAVESSP 840
 781 EPRTGLSNBRAMGRFDSVTGPNGACVPAFCDDQSPDAYSAVYLAISLAGEBLOAVESSP 840
 Db 781 EPRTGLSNBRAMGRFDSVTGPNGACVPAFCDDQSPDAYSAVYLAISLAGEBLOAVESSP 840

QY 841 KFPNNAIGVOPPLINKNTNRTDHEPRVKKTAFOISMAKPRPSAENSNGPIYAFENLR 900
 841 KFPNNAIGVOPPLINKNTNRTDHEPRVKKTAFOISMAKPRPSAENSNGPIYAFENLR 900
 Db 841 KFPNNAIGVOPPLINKNTNRTDHEPRVKKTAFOISMAKPRPSAENSNGPIYAFENLR 900

QY 901 ACEBAPSAHFRFYQEGRYDNTVPFNEDDPMSTEDYLAWPMPMEPRACYIKVKI 960
 901 ACEBAPSAHFRFYQEGRYDNTVPFNEDDPMSTEDYLAWPMPMEPRACYIKVKI 960
 Db 901 ACEBAPSAHFRFYQEGRYDNTVPFNEDDPMSTEDYLAWPMPMEPRACYIKVKI 960

QY 961 VGPLEVNVSRNNGGTHRTVYGLKLRDVRSTRDQPNVSAACLEFKSGMLYDDDRY 1020
 961 VGPLEVNVSRNNGGTHRTVYGLKLRDVRSTRDQPNVSAACLEFKSGMLYDDDRY 1020
 Db 961 VGPLEVNVSRNNGGTHRTVYGLKLRDVRSTRDQPNVSAACLEFKSGMLYDDDRY 1020

QY 1021 DRTLKVIIPQSCRRASVNMPLHEYLVLNMLPLAVNNDTSEYMLADPLIGHNVGYTVT 1080
 1021 DRTLKVIIPQSCRRASVNMPLHEYLVLNMLPLAVNNDTSEYMLADPLIGHNVGYTVT 1080
 Db 1021 DRTLKVIIPQSCRRASVNMPLHEYLVLNMLPLAVNNDTSEYMLADPLIGHNVGYTVT 1080

QY 1081 DQBPRTAKETALGRCPFGTSDGSSRIKSNVGVALTFCNCEYRGVGRASAOYIQTSPAQS 1140
 1081 DQBPRTAKETALGRCPFGTSDGSSRIKSNVGVALTFCNCEYRGVGRASAOYIQTSPAQS 1140
 Db 1081 DQBPRTAKETALGRCPFGTSDGSSRIKSNVGVALTFCNCEYRGVGRASAOYIQTSPAQS 1140

QY 1141 PAAGTVGRVPSRROQASRGQOSGVASLPRPRAOOLIN 1184
 1141 PAAGTVGRVPSRROQASRGQOSGVASLPRPRAOOLIN 1184
 Db 1141 PAAGTVGRVPSRROQASRGQOSGVASLPRPRAOOLIN 1184

RESULT 2
 US-08-996-083-3
 ; Sequence 3, Application US/08996083A
 ; Patent No. 6124095
 ; GENERAL INFORMATION:
 ; APPLICANT: Magna, Holly
 ; APPLICANT: Schaffer, Paul
 ; APPLICANT: Lawton, Michael
 ; APPLICANT: Yocum, Sue
 ; APPLICANT: Mitchell, Peter
 ; APPLICANT: Hutchinson, Nancy
 ; APPLICANT: Murty, Lynn E.
 ; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2
 ; FILE REFERENCE: PF-0420 US
 ; CURRENT APPLICATION NUMBER: US/08/996,083A
 ; CURRENT FILING DATE: 1997-12-22
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 1184
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; OTHER INFORMATION: Incyte Clone No. 6124095: 422069
 ; PUBLICATION INFORMATION:
 US-08-996-083-3

Query Match 99.7%; Score 6297; DB 3; Length 1184;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1181; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGTGKAVSFLEVTSLVGRQTMLTQSVRRVOPGKKNPSIFAKPADTLESFGEWTTLF 60
 1 MGTGKAVSFLEVTSLVGRQTMLTQSVRRVOPGKKNPSIFAKPADTLESFGEWTTLF 60
 Db 1 MGTGKAVSFLEVTSLVGRQTMLTQSVRRVOPGKKNPSIFAKPADTLESFGEWTTLF 60

QY 61 NIDYGGKGYERLDAIRFYGGDRCARPLRLRARTDWTDPAGSTGVVHSGPREGFWCL 120
 61 NIDYGGKGYERLDAIRFYGGDRCARPLRLRARTDWTDPAGSTGVVHSGPREGFWCL 120
 Db 61 NIDYGGKGYERLDAIRFYGGDRCARPLRLRARTDWTDPAGSTGVVHSGPREGFWCL 120

QY 121 NRORPQONCSNTVRLCPGSLRRDTERIWSBPSWMSKSAACGQTVQTRTRICLAE 180
 121 NRORPQONCSNTVRLCPGSLRRDTERIWSBPSWMSKSAACGQTVQTRTRICLAE 180
 Db 121 NRORPQONCSNTVRLCPGSLRRDTERIWSBPSWMSKSAACGQTVQTRTRICLAE 180

QY 181 MVSICSEASEGQHCKQDCTACDLTCPMGOVNADCDACACODPMLHGAVSLPGAPASG 240
 181 MVSICSEASEGQHCKQDCTACDLTCPMGOVNADCDACACODPMLHGAVSLPGAPASG 240
 Db 181 MVSICSEASEGQHCKQDCTACDLTCPMGOVNADCDACACODPMLHGAVSLPGAPASG 240

QY 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKYKFAPIYLTMTKTSIKAATIK 300
 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKYKFAPIYLTMTKTSIKAATIK 300
 Db 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKYKFAPIYLTMTKTSIKAATIK 300

QY 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCATGKPRPDXYFWYHNDTLDPISLYKHESK 360
 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCATGKPRPDXYFWYHNDTLDPISLYKHESK 360
 Db 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCATGKPRPDXYFWYHNDTLDPISLYKHESK 360

QY 361 LVLRKLDQOHAGEFYCKAQSADAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDFQON 420
 361 LVLRKLDQOHAGEFYCKAQSADAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDFQON 420
 Db 361 LVLRKLDQOHAGEFYCKAQSADAGAVKSKVAQLIVTASDETPCNPVPESYLIRLPHDFQON 420

```

OY 421 ATNSFYDVGRCPVKTACAGQDNGIRCDAYONCCGISKEEREIOCSGYLPTKVAKBC 480
DB 421 ATNSFYDVGRCPVKTACAGQDNGIRCDAYONCCGISKEEREIOCSGYLPTKVAKBC 480
OY 481 SCORCTETRSIVRGVSAADNGEPMFGHVYMGNSRVSMTGKGTFTLVHPDTERLVLT 540
DB 481 SCORCTETRSIVRGVSAADNGEPMFGHVYMGNSRVSMTGKGTFTLVHPDTERLVLT 540
OY 541 FVDRLQKFPVNTKVLFPNKKGSVAVHEIKMLRKKEPITILEAMETNIIPGEEVGEPMAE 600
DB 541 FVDRLQKFPVNTKVLFPNKKGSVAVHEIKMLRKKEPITILEAMETNIIPGEEVGEPMAE 600
OY 541 FVDRLQKFPVNTKVLFPNKKGSVAVHEIKMLRKKEPITILEAMETNIIPGEEVGEPMAE 600
DB 541 FVDRLQKFPVNTKVLFPNKKGSVAVHEIKMLRKKEPITILEAMETNIIPGEEVGEPMAE 600
OY 601 LEIPRSRFRQNGEPYIGVKASVTFPLDPNISTATAOTDINFIDEEDTPELPTRYGME 660
DB 601 LEIPRSRFRQNGEPYIGVKASVTFPLDPNISTATAOTDINFIDEEDTPELPTRYGME 660
OY 661 SVDFRDEVTSEPLNAGKVYVHLDSQVKKMPEHISTVKLMSLNDPTGLMEEGDFKEENOR 720
DB 661 SVDFRDEVTSEPLNAGKVYVHLDSQVKKMPEHISTVKLMSLNDPTGLMEEGDFKEENOR 720
OY 721 RKKRDRFTLVGNLEIRERLFLNDVBPESRCFVKYAVRSEERFLPSBOIQGVISVIML 780
DB 721 RKKRDRFTLVGNLEIRERLFLNDVBPESRCFVKYAVRSEERFLPSBOIQGVISVIML 780
OY 781 EBRTEGLSNPRAMGRPDSVITGPNGACVAFCDOSPDAYSAYVLASLAGEELQAVESSP 840
DB 781 EBRTEGLSNPRAMGRPDSVITGPNGACVAFCDOSPDAYSAYVLASLAGEELQAVESSP 840
OY 841 KENPAIGVPODYLNKLNRTDHEDPVKYKTAFOISMAKPPRNSAESNGPIYAFENLR 900
DB 841 KENPAIGVPODYLNKLNRTDHEDPVKYKTAFOISMAKPPRNSAESNGPIYAFENLR 900
OY 901 ACEEAPPSAAHRRFQIIEEDRDYNTVPENEDDPMSTEDYLAAMPKPEFRACITKVKI 960
DB 901 ACEEAPPSAAHRRFQIIEEDRDYNTVPENEDDPMSTEDYLAAMPKPEFRACITKVKI 960
OY 961 VGPLEVNRSMNGGTHRTYVKLGIRDVRSRTRDROPNVSAACLEFGCSGLYDODAV 1020
DB 961 VGPLEVNRSMNGGTHRTYVKLGIRDVRSRTRDROPNVSAACLEFGCSGLYDODAV 1020
OY 1021 DDTLVKVIPOGSCRASVNPMLHEYLVLNHLPLAVNNDTSEYTMALPLDLGHNHYGYTYT 1080
DB 1021 DDTLVKVIPOGSCRASVNPMLHEYLVLNHLPLAVNNDTSEYTMALPLDLGHNHYGYTYT 1080
OY 1081 DDDPTAKETIAGRCFDTSDGSSSRIMKSNVVALTFNCVEROVGRQSAFYLOSTPAOS 1140
DB 1081 DDDPTAKETIAGRCFDTSDGSSSRIMKSNVVALTFNCVEROVGRQSAFYLOSTPAOS 1140
OY 1141 PAAGTVQGRVPSRRQORASRGOROSGVVASLRFPVAAOOPILIN 1184
DB 1141 PAAGTVQGRVPSRRQORASRGOROSGVVASLRFPVAAOOPILIN 1184

```

```

RESULT 3
US-08-996-083-1
Sequence 1, Application US/08996083A

```

```

GENERAL INFORMATION:
APPLICANT: Magna, Holly
APPLICANT: Schaffer, Paul
APPLICANT: Lawton, Michael
APPLICANT: Yocum, Sue
APPLICANT: Mitchell, Peter
APPLICANT: Hutchinson, Nancy
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2
FILE REFERENCE: PF-0420 US
CURRENT APPLICATION NUMBER: US/08/996, 083A
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1

```

```

; LENGTH: 1156
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte Clone No. 6124095; 1388013
; PUBLICATION INFORMATION:
US-08-996-083-1

Query Match 48.9%; Score 3091.5; DB 3; Length 1156;
Best Local Similarity 50.3%; Pred. No. 5,3e-275;
Matches 584; Conservative 178; Mismatches 359; Indels 39; Gaps 12;

OY 24 TMLTOSVRVQEGKKNPSIFARADTLESBGEWTTLFNIDYPOGKDDYERLDAIRRYGD 83
DB 32 TALGERRSVYTGQPSPAL-----EDMEASEWTSWFNVDHPDGDDEFSLAIRPYGP 86
OY 84 -RYCARPLREARTTMTWTSQGVVHGSPRGFCMLNREORPGONCSNYTRFLCPCPG 142
DB 87 ARVCPRLLEARTTDMALPSANGVERVHLNPTRGFCMLNREORPGRCNSYHVRFCP-- 144
OY 143 SLRDRTERIWSFSPWSKCSAAGOTGVOTRRICLAENVSLCSEASEGOMCMGDCYA 202
DB 145 -----LEASMGAMGPMWGSOGSG- -PGRRLRRHCHPSPGADCGRPLEAKVRRPCG 198
OY 203 CDLTGCMGOVNADCDKCMODEMLHGAVSLPGAPASGAIIYLTTPKLTOTDSDBGF 262
DB 199 CSL-----DICEPDDHLLGSVYTPSQDPLLAGRVSLRD- -PGTVATSDAHGTF 246
OY 263 RIPLGCPDGKSLIKTKYKF-APIVLTMPKTSLKATTAEFVBARTEPYVMNPETKARR 321
DB 247 RVPGVCAADRANIRAOQMDGFSAGEAOANGSISVYTTILD- -KLEKPLVKKHESRYVE 304
OY 322 AGQSVSLCCATGKPRPDYFYWHNDTLDPSSLYKHESKLVYRKLQOHQAGEYFCKAOSD 381
DB 305 AGQNTVECCKASGTPMKRYSMFNHGTLLDRAHNGGAHLELRGLRPGDAGIYHCKAMNE 364
OY 382 AGAVKSKVAVQLYVTSADDEPCNPVPESYILRLPHCFQONATNSFFYDVGRCPVKTACAGQ 441
DB 365 AGAVRSGLARTLVLAGQPCDPRPREYLIKLPEDGQGSQSGAYLDVGICPTTRCPSLA 424
OY 442 DNGIRCDAYONCCGISKEEREIOCSGYLPTKVAKECSORCTETRSIVRGVSAAD 501
DB 425 GSSPRCGDASSRCCSVKRLERNEHICPGYVLPKYVAECGCOCLPRLKLVKRGVAAAS 484
OY 502 GEMPRGHHVYMGNSRVSMTGKGTFTLVHPDTERLVTFVDRLOKFEVNTTKVLPFNKKG 561
DB 485 GEPLRFAARILLGOEPIGFTAYOGDFTIEVPPSTQRLVYTFVDPSSGFMADAVRLPDPDGG 544
OY 562 SAVFEIKMLRKKEPITILEAMETNIIPGEEVGEPMAELETPSRSFYQNGEPTYGKVK 621
DB 545 AGYHHEVKAMRKKAIVILHTSOSNTIPLGELEDEAVLGEIVLPVSGAFRADGPKYSGPVE 604
OY 622 ASVTFLDPNISTATAOJDNLNFINDEGDTPLRTYGMFSVPFRDVTSEPLNAGKVYKH 681
DB 605 ARVTEVDPDRDLTSASASPSDLFRVDSDEGLAPLRITGYMFSVDLRADGSAEQLOVGVAAVR 664
OY 682 LDSTOVKMPHEHISTVKLMSLNDPTGLMEEGDFKEENOR- -RKKRDRFTLVGNLEIRER 739
DB 665 VASQIHPMGVHAEKLTMSLNPETGLMEESGRRGSGSPRYRREVRFLVGNVEIRER 724
OY 740 RLFLNLDVBPESRCFVKYAVRSEERFLPSBOIQGVISVYNLEPRGFLSLNPRAMGRFDSV 799
DB 725 RLFLNLDVBPESRCFVKYAVRSEERFLPSBOIQGVISVYNLEPRGFLSLNPRAMGRFDSV 784
OY 800 ITGPNACVAPACDDQSPAYSAYVLASLAGEELQAVESSPKKNPAITGVPODYLNKLVY 859
DB 785 VTGPNACVAPACDDQSPAYSAYVLASLAGEELQAVESSPKKNPAITGVPODYLNKLVY 844
OY 860 RTDHEDPVKYKTAFOISMAKPPRNSAESNGPIYAFENLRACEEAPPSAAHRRFYQIG 919
DB 845 RTDHDPAKFRNGRFINLAKPRPDGPAEANGVYVWRSRLREGOGAPVYASHRFAVEA 904

```

```

OY 920 DRYVNVLPFNEDDPMSTEDYLLAMMPKPMFRACYIKVKIVGLEVNVSRNMGGTHRR 979
Db 905 DKRENVVPEFREGPASTBTGULLAMWNPQEFRCFLKVKIQGQELVMWSHNAAGSHPR 964
OY 980 TVGKLYGIRDVSTRDRDQPNVSAACLEFKCSGMLYDQDRVDRTLVKVIFQSCCRASVN 1033
Db 965 TRGQVLYGLRDARSVRDEPERPETSAAACEFKCSGMLFDQORVDRTLYV IMPQSCCRVAVN 1022
OY 1040 PMHELYLVNHLPLAVNNDTSEYTMALPLDPLGHNVGYIYVTDOPPRAKETALGRCPDGT 109
Db 1025 GLRKYLTRHPPEYPAEDPFAFSLAPLADPLGHNVGYIYVTDOSPRLAKETALGRCPDGS 108
OY 1100 SDGSSRIKMSNVGVALTEFCVEROVGRQSAFOYLQSTPAOSPAAGTVQGRVPSRQORAS 115
Db 1085 SDGFSREKMAADAGTAVTFQCREPPAGRPSLTQRLLESPPA--TALGDIR-----REMSSEA 1133
OY 1160 RGGORQSGVASLRFPRVAQ 1179
Db 1138 QAOARASGPLETRRR-GRVRQ 1156

OUT 4
US-09-429-516-1
; Sequence 1, Application US/09429516
; Patent No. 6251389
; GENERAL INFORMATION:
; APPLICANT: Magna, Holly
; APPLICANT: Schaffer, Paul
; APPLICANT: Lawton, Michael
; APPLICANT: Vocum, Sue
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOS
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/429, 516
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/996,083
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0420 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1156 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SATPFI008
; CLONE: 1388013,

```

US-09-429-516-1

Query Match	48.9%	Score 3091.5	DB 4	Length 1156
Best Local Similarity	50.3%	Fred No. 5.3e-275		
Matches 584	Conservative 178	Mismatches 359	Indels 39	Gaps 12
24	TMLTOSVRVQGGKKNISIFAKPADTLESPEBMTTLFNIDYGGKGVDELRLAIFRYGD	83		
Db	32 TALGLERSVYIGQSPAL-----EDWEASMTSMFNVDHGGGDPESLAIIRFYGP	86		
QY	84 -RYCARPLRLARTTDMTPAGSTGOVYHGSPEEGWCMLNREORPCQONSNYIVRELCPG	142		
Db	87 AAVCPRLALERTTDMALPSAVGRIWHLNLPGRGFCMLNREORPCRCNSYHVRCP--	144		
QY	143 SLRDRTERIWSWSPMSKCSACGGTGVOTPRICIAEWSYSCSBAESEGQCMQDCTA	202		
Db	145 -----LEASWGMGWMGPGCSGCG-PGRRLRRKHCHPSPAIGDACPGRLPENAQCVRPRCG	198		
QY	203 CDLTCPMGQVNAADCACMCQDFMLHGAVSLPGGAPASGAAYLLTKPKLLTQDSDGRF	262		
Db	199 CSL-----DTCECPDHLILGVSVTPEGQPLIGARVSLRQ-PQTVATSDAHGTF	246		
QY	263 RIPGCLDPGKSLTKTKTKYF-APYLTMKTSYKATITAEVRAETPMVANNPEFKAR	321		
Db	247 RVPGACADSRATIRQMDGFSAGEAOQANGISVYTIILD--KLEKPLVNHPSERVHE	304		
QY	322 AGOVSILCCKATGKPRPDYFYWHDTLLDPSLYKHESKLVLRKLQHQAGEYFCKAOSD	381		
Db	305 AGQNTVFCCKASGTPMPKKYSWFHNGHTLLDRANHGGAHLRLGLRPDQAGIYHCKANNE	364		
QY	382 AGAVSKYAQLIVTASDEFPCNPVSESTYLRPHCFQMATSFYDYDRCRVKTCAGQO	441		
Db	365 AGAVSSGTAARLVLPAGQADCPRPREYLIKLPEDCGQGSPPAYLDVGLCDPTRCPSIA	424		
QY	442 DNGICRDVAQCCISIKTEEREIOCSGTTLETYKAKECSQCRCTETRSIVGRVSAADN	501		
Db	425 GSPSPCGAASSCCSVRRLEIRREHOCGVLPVYKVAECCQCKLPRLGRLVGRVVAADS	484		
QY	502 GEPARFGHYMGNSKRYSMTGKGYFTLHVPOTERLYTLFVDRLOKEVYTTKVLPPNKG	561		
Db	485 GEPRLRRLITQOEPIGFAYOGDFTEIEXPSTORLVTFVPEGSEFMQAVRLPLPDPG	544		
QY	562 SAVFHEIKLRKEPITTEAMETNIIPLCEVVGEDMMALETPSRFSFYRQNGEPIYGYKX	621		
Db	545 AGVYHEVKAARKKAPVILITLTSQSNITPLGELDEARLGELVLPQSGFRADCKPYGPE	604		
QY	622 ASVTELDPRNISTATTAQOTLTLNFINDEGDTPLRTYGMFSVDFRDEYISEPLNACKVKYH	681		
Db	605 ARVTEVDPDLTSAASAPSDLRFVSDGDELABLRITYGMSVLDLRAFCSAEOLQVGVAAVR	664		
QY	682 LUSTOYKMEHEHSTYKWLMLNPDGTGMEEGGPKFNQK--RHKREDRFLVGNLEIRER	739		
Db	665 VVASQTHMHGHAELKWLMLNPEITGMEESGFRREGSSGPRVRRERERVLGNVETIRER	724		
QY	740 RLENLDPESRRCFYKVAARYRSERELPSBOJOGVYLSVNLBPRIGFLSLNPRAMGRFDSV	799		
Db	725 RLENLDPERRRCFYKVAARYANDKFTSPBOYGVVVYTLVNLBPAGFSANPRAMGRFDSA	784		
QY	800 ITGPNAGCYPAFCDDOSPAYISAYVLASLAGELQAVESSPKFNPALIVGPQPYLKNLNY	859		
Db	785 VYGPNGACLPACCDADRPAYATALYATLIGBELPEAPSLPRLPATVGVTOPLYDRLRGY	844		
QY	860 RRTDHEDDPVKKTARFQISNAKPRPNASAEESNGYIAFYEMLRACEEAPSAAHFRFOIGY	919		
Db	845 KRTHDDHDPFKKRGFTINLAKPRPGDPAFANGFYVYWRSLRECQCAPYTAHSFRFRKYVA	904		
QY	920 DRYDYNTVPFNDDPSMTEDYLAMWPKMEERACYYKAYIGPTEVYNSRMNGSTHRR	979		
Db	905 DKYEYVNVVPEFRGTPASMTGDLLAMWPNQPEFRACFLKVIQGPQGYWYRSNAGGSBPR	964		
QY	980 IYKGLGATDVYSTRRDQPNVSAACLEKCGSMALYDQDRIRTVLKVIPQSGCRASVYN	1039		

Db 965 TRGOLYGLRDARSVDPERPCTSAACVEFKCSGMLFDOROVDRTLVLTIMPOGSCRRVAVN 1024
Qy 1040 PMLHELYLVNHLPLAVNNDTSEYTMPLAPDLGHNGYITVTDODPRFAKETAIGRCFCGT 1099
Db 1025 GLNDYLTFRHPPVPAEDPAFSLAPLDPLGHNGYITVTDOSRLAKETAIGRCFCGS 1084
Qy 1100 SDGSSRIKSNVGVALTFCNVEROVROSARFOYLOSTPAOSPAAGTYOGRVPSRQORAS 1159
Db 1085 SDGSRREKADAGTAVTFQCCEPPAGRPSLFRLESPA--TALGDIR-----REMSRA 1137
Qy 1160 RGGOROSGVASLREPRVAQ 1179
Db 1138 QAOARASGRLTRR-GRVRO 1156

RESULT 5
US-09-429-516-3
Sequence 3, Application US/09429516
Patent No. 6251389
GENERAL INFORMATION:
APPLICANT: Magna, Holly
APPLICANT: Schafter, Paul
APPLICANT: Lawton, Michael
APPLICANT: Yocum, Sue
APPLICANT: Mitchell, Peter
APPLICANT: Hutchinson, Nancy
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOS
NUMBER OF SEQUENCES: 3
TITLE OF INVENTION: PHOYDROLASE-2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/429,516
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/996,083
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: P-0420 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1156 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SATP1002
CLONE: 422069
US-09-429-516-3

Query Match 48.98; Score 3091.5; DB 4; Length 1156;
Best Local Similarity 50.38; Pred. No. 5.3e-275;
Matches 584; Conservative 178; Mismatches 359; Indels 39; Gaps 12;

Qy 24 TMLTQSVRRVOPGKKNPISIFAKPADLTSPGEMTTLFNIDYPGKGKDYERLDAIRFYCD 83
Db 32 TALGERSVYTTGQSPAL-----EDMEASGEMTSMFVNDHGGGDESLAIRFIYGP 86
Qy 84 -RVCARPLRLERATTDWTPAGSTGOVHGSPREGFCMLNREORPGONCSNTYVFLCPLP 142
Db 87 ARVCPRLALARTDMLPSAVGERVHLNPTRGFWCLNREQPRGRCSNHYHFRCP-- 144
Qy 143 SLRRTERTINSPWMSWMSCSAAGCOTGVOTRTRICLAEMVSLCSASREGCHKODCTA 202
Db 145 -----LEASWAGMGPWGPCSSCG--PGRRLRRHRCPSAGACPRPLEAOKCVAPRCPG 198
Qy 203 CDITCPMGQVANADCACQCOFPMILGAVSLPGCAPASAAIYLLTKTKLLTOTSDRF 262
Db 199 CSL-----DTCECPHILLAGSVYTPSGOPLDARVSLRQ--PGYATVSAHGTF 246
Qy 263 RIPGLCPDGKSLIKITKYK--APIVLTMPKTSLKATTKAEFVRAETPYVMNPETKARR 321
Db 247 RVPVCADSRANIRAOQMGFSAGEAQANGSISVVTILD--KLEKPYLVKHPESRYRE 304
Qy 322 AGQSVSLCCKATGKPRDPKTYMINDTLDPSLYKHESKLVLRKIQHQAQAEYFCKAOSD 381
Db 305 AGQNVTFCCCKASGTMPMKYSWFMNGTLLDRRAHGYAHLEIRLPDQAQGIYHCKANNE 364
Qy 382 AGAVSKVAOLIVTASDETPCNPVPESTLIRLPHDCFQMNATNSFYDVGRCPPVKCAQO 441
Db 365 AGAVASGTARLTLVLAGGPACDPRETYLILPECCGPGSGPAYLDGLCPTDRCPSLA 424
Qy 442 DNGINCRDVAONCCGISKTEREIOCSGYTLPTKYAKESQOCTETRSIVRGVSAADN 501
Db 425 GSPRCGDASSRCSVRRLREIRHCPGYLVKVAACGCGQKCLPRGLVRGRVVAADS 484
Qy 502 GEPARFGHYVNGNSRVSMYTKGTTLVPODTELVLTFDRLQKFPVNTTKVLPFNKG 561
Db 485 GEPLRFARLILGOEPIGTAVOGDETIYVPPSTORLTVFDPSCGEFADAVRVLFPDRG 544
Qy 562 SAVFEIKMLRKEPITLEAMETNIIPGEVVGEDPMALFISPSFYRONCEPIYGVK 621
Db 545 AGVYHEVAMARKKAPVILHTSOSNTIPIGLEDENAPLELVLPSCAFRADCKPISGYVE 604
Qy 622 ASVTFDLDRNISTATAOTDLNFINDEGDTPELRTYMGESVDFRDEVTSEPLNACKVYH 681
Db 605 ARVFVDPDRDLTSAASASDRLFVDSDELAPLRTYMGFSVDLRAFGSAEOLQVGPVAVR 664
Qy 682 LDSTOVKAPHEIISTYKLSLNDPTGLMEBGDFEKNOR--RNKREDTFLVGLNETIER 739
Db 665 VAAQIHPGHVGEALKLSLNDPTGLMEBEGFRREGSSGPRVREERVFVGVNVEIRER 724
Qy 740 RLENLDVPESTRCEFYKVAYSERFLPSEOLQGVYISYINLEPRGFLSNPRAMGRPSV 799
Db 725 RLFLNDVPERRRCFVKVAIYANDKTPSEQVEGVYVTLVNLFPAGFSANPRAMGRFSA 784
Qy 800 ITTPNGACVPACDQSPDAYSAVYVLAISAGELOAVESSPKFNMAIGVPPYLNLNY 859
Db 785 VTGPNAGCLPAFCADRDADATYATLATLGGEELEPAPSLPRPLATVAGVGYDRLRGY 844
Qy 860 RRTHEDEPRVAKTAQIOMAKRPNRPSAEBSNGPIYAFENLRACEAPSAHFRYQLEG 919
Db 845 RRTDHDPAFRKNGRINLARPRPDPAEANGPVVPMBSLRCCOGAPYASHFRARVEA 904
Qy 920 DRYDNTVPFNDDPMSTEDVLIAMPPKPMERFRACYIKYIGPLEVNVRSNMGGTIRR 979
Db 905 DKLEYNVYFPREGTPTASMTGDLIAMPNPQERFRACFLKVKILOGPOEYVNRSHNAGCSHPR 964
Qy 980 TVGKLYGIRDVSTRDRQDPNVSAACLEFKCSGMLYDODRVDRTLVKYIPGOSCRASVN 1039
Db 965 TRGOLYGLRDARSVDPERPCTSAACVEFKCSGMLFDOROVDRTLVLTIMPOGSCRRVAVN 1024
Qy 1040 PMLHELYLVNHLPLAVNNDTSEYTMPLAPDLGHNGYITVTDODPRFAKETAIGRCFCGT 1099
Db 1025 GLNDYLTFRHPPVPAEDPAFSLAPLDPLGHNGYITVTDOSRLAKETAIGRCFCGS 1084
Qy 1100 SDGSSRIKSNVGVALTFCNVEROVROSARFOYLOSTPAOSPAAGTYOGRVPSRQORAS 1159

DB 1085 SDFSTHEMADAGTATVPCORREPPAGRPISLFOFLLESPPA--TALGDIR-----REMSEMA 1137
 QY 1160 RGSOROSGVVASLRFPRVAQ 1179
 DB 1138 QAQARASGLPTRRR-GRVRO 1156

RESULT 6

US-08-954-333-7
 ; Sequence 7, Application US/08954333
 ; Patent No. 5986080
 ; GENERAL INFORMATION:
 ; APPLICANT: Ikuko Masuda, et al.
 ; TITLE OF INVENTION: CLONED NUCLEOTIDE PYROPHOSPHOHYDROLASE AND
 ; TITLE OF INVENTION: USED THEREOF
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Quarles & Brady
 ; STREET: 411 East Wisconsin Avenue
 ; CITY: Milwaukee
 ; STATE: Wisconsin
 ; COUNTRY: U.S.A.
 ; ZIP: 53202-4497
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/954,333
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ryser, David G.
 ; REGISTRATION NUMBER: 36,407
 ; REFERENCE/DOCKET NUMBER: 650053, 91070
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (414) 277-5717
 ; TELEFAX: (414) 271-3552
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 599 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-954-333-7

Query Match 45.9%; Score 2898.5; DB 2; Length 599;
 Best Local Similarity 91.1%; Pred. No. 9,6e-258;
 Matches 544; Conservative 23; Mismatches 29; Indels 1; Gaps 1;

QY 588 PLGEVVGEDPMALIEIPSSRYRQNGEPYIGVKASVTFPLDPNNISTATATQDTLNFIND 647
 DB 4 PLGDMGDEPMGLIEIPSKSFYRQNGEPYIGVKASVTFPLDPNNISTATATQSDNFIND 63
 QY 648 EGGTFPLRTYGMFSVDFPREVISEPLNACKVKVHLDSTQVKMPEHISTKLSLNDPTGL 707
 DB 64 EGGTFPLRTYGMFSVDFPREVISEPLNACKVKVHLDSTQVKMPEHISTKLSLNDPTGL 123
 QY 708 WEDEGDFKFNORRNKREDTFLVNLIEIRERLFLNDVPSRRCFVKVAVRSERFLPS 767
 DB 124 WEDEGDFKFNORRNKREDTFLVNLIEIRERLFLNDVPSRRCFVKVAVRSERFLPS 183
 QY 768 EGIQGVIVSINLEPTGFLSNRANGRPDSVTGNGACVPAFCDDOSPDAYSAYVLAS 827
 DB 184 EGIQGVIVSINLEPTGFLSNRANGRPDSVTGNGACVPAFCDDOSPDAYSAYVLAS 243
 QY 828 LAGEELQAVSSPFFNFNAIGVPOPYLNKLINERTDHEDPVAKTAFOISMARPRNSAE 887
 DB 244 LAGEELQAVSSPFFNFNAIGVPOPYLNKLINERTDHEDPVAKTAFOISMARPRNSAE 303

QY 888 ESNQPIYAFENLRACEAPPSAAHPRFYQIEGRDYDNTVPENEDDDMSWTEDYLAMPK 947
 DB 304 ESNQPIYAFENLRACEAPPSAAHPRFYQIEGRDYDNTVPENEDDDMSWTEDYLAMPK 363
 QY 948 PMEFKACIYKVIQPLEVNVSRNMGTHRTVTKLYGIRDVSRDRODPNVSAACLE 1007
 DB 364 PMEFKACIYKVIQPLEVNVSRNMGTHRTVTKLYGIRDVSRDRODPNVSAACLE 423
 QY 1008 FKCSGMLDQDVRDRTLVKVIPOGSCRRASVNPMLHEVLVNLPLAVNPDSEYTLAPL 1067
 DB 424 FKCSGMLDQDVRDRTLVKVIPOGSCRRASVNPMLHEVLVNLPLAVNPDSEYTLAPL 483
 QY 1068 DPLGHNHYGIYTVDDPRTAKELALGRCEDTSDGSSRIMKSNVGVALLFNCVEROYGRQ 1127
 DB 484 DPLGHNHYGIYTVDDPRTAKELALGRCEDTSDGSSRIMKSNVGVALLFNCVEROYGRQ 543
 QY 1128 SAFQYIQTSTPAOSPAAGTYGGRVPSRROORASRGOROSGVVASLRFPRVAQOPLN 1184
 DB 544 SAFQYIQTSTPAOSPAAGTYGGRVPSRROORASRGOROSGVVASLRFPRVAQOPLN 599

RESULT 7

US-08-313-288B-20
 ; Sequence 20, Application US/08313288B
 ; Patent No. 5750502
 ; GENERAL INFORMATION:
 ; APPLICANT: Jessell, Thomas M. and Avihu Klar
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
 ; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/313,288B
 ; FILING DATE: January 5, 1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0526
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1170 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-313-288B-20

Query Match 2.5%; Score 157; DB 1; Length 1170;
 Best Local Similarity 20.1%; Pred. No. 5.2e-05;
 Matches 111; Conservative 37; Mismatches 150; Indels 254; Gaps 27;

QY 10 SLTVLEWISVLGROMLQSVARVOPGRKNPSIFAKPADTLESP-----GEW 56
 DB 278 SSMVLELRKRLRTVTTLDOSIRKYTEENK-----ELANEELRRRLPLCYHNGVQYRNDEW 331

```

OY 57 TT-----LFNIDYPGKGDEYERLDAIRFYGGDRVCARPLRL 92
DB 332 TVDSCTECHCONSVYICKKVSCLPIMPCSNATVPDGE-----CCRCWMS 375
OY 93 EARTDWTAPAGSTGOVYHGSPPREGFW-----CLNREORPGONCSNYTVRFLCPPGSLRR 146
DB 376 DSADGWSWSPWSE-----WTSCSTSCGNGIOGRGSCDILNRR--CEGSSVQT 420
OY 147 DHERI-----WSPWSPWKSCKSAACGOTGVOTRTRICLAEMVSL-----CSEASE 190
DB 421 RCHHIOECCKRFRKODGWSHWSWSSSVYTCGD--GVITRIRLCNSPSPOMNKPCEGEAR 479
OY 191 ESOHMGODC-----TACDLTC-----PMGOVNA-DC-----D 217
DB 480 ETKACKKADACPINGGWSWSPWDICSVYTCGGVOKRSRLCNPTPOFGKDCVGVTEWQ 539
OY 218 ACMODFMLHGAVSLPGCAPASGAAYLLTKPKLLTQTDSDGRPRIPGLCPDGSILKI 277
DB 540 ICKKODCPIDGCLSNPCFAGVK-----CTSTP-----GSMWC--GACPPGYS----- 580
OY 278 TVKRAPIVLTMPKTSIKATIKAEFVRAETPYMMNPETKARRAGOSVSLCKKATGRPR 337
DB 581 -----GNGIOCTVDVECKEV 595
OY 338 PDKYFWYHNDLLDPSLYKHESKLYLRKLOHQAGEYFCFKAOSDAG----- 383
DB 596 PDACE--NNH-----GEHRE--NTDPOYCNLPCPPRTGSG 628
OY 384 ----AVKSKVAOLIVTASDETPCNVPESYLRIRPHDGFONATNSF---YVD--VGRCPVK 435
DB 629 PFGQVEVHAHTANKOVC---KPRNPCTDG---THDCNKNAKCNVLI GHYSDPMYRCECK 679
OY 436 TCAGOODNGIRG 447
DB 680 --PGTAGNGITC 689

```

RESULT 8

US-08-918-914-4

Sequence 4, Application US/08918914

Patent No. 5876963

GENERAL INFORMATION:

APPLICANT: Mitchell, Peter

APPLICANT: Hutchinson, Nancy

APPLICANT: Lawton, Michael

APPLICANT: Magna, Holly

APPLICANT: Murry, Lynn E.

TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/918,914

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0369

TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 788 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1070094
US-08-918-914-4

```

Query Match 2.3%; Score 147.5; DB 2; Length 788;
 Best Local Similarity 27.2%; Pred. No. 0.00019;
 Matches 55; Conservative 15; Mismatches 65; Indels 67; Gaps 11;

```

OY 46 PADTLESPEWMTTLFNIDYPGKGDEYERLDAIRF-----YYGDRVCARPLRL 93
DB 501 PCQTFSEWCEWST---CSASCSGGORER---TRFCHLGTNRCEGKDYSEOC SAGCPBW 554
OY 94 ARTFTWTPAGST--GOVYHGSPPREGFWCL-----NREOR--PGONCSNYTVRFL 138
DB 555 SQWEDMGQCSVYTCGGGVAVRQRT---CLGGVFGDHLGQGPKEGRACDGGPCS----- 604
OY 139 CPPGSLRDTERTIWSFWSWKSCKSAACGOTGVOTRTRICLAEMVSLCSFASSEGOHMCQ 198
DB 605 -----LWSPWQDMSVCSASCG--SGMKRRORVC--OFGTDCOGPNESOFQVCP 649
OY 199 DC-----TACDLTCPMQ 211
DB 650 PCAEWTEMCWMSGSSKCGPGQ 671

```

RESULT 9

US-08-348-006B-5

Sequence 5, Application US/08348006B

Patent No. 5658756

GENERAL INFORMATION:

APPLICANT: RODAN, GIDEON A.

APPLICANT: SCHMIDT, AZRIEL

APPLICANT: RUTLEDGE, SU JANE

TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: J. MARK HAND

STREET: 126 E. LINCOLN AVE., P.O. BOX 2000

CITY: RAHWAY

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07065-0900

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC Compatible

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/348,006B

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/122,032

FILING DATE: 14-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: HAND, J. MARK

REGISTRATION NUMBER: 36,545

REFERENCE/DOCKET NUMBER: 189921A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 908-594-3905

TELEFAX: 908-594-4720

; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1911 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-348-006B-5

Query Match 2.3%; Score 144; DB 1; Length 1911;
 Best Local Similarity 19.5%; Pred. No. 0.0019;
 Matches 186; Conservative 122; Mismatches 362; Indels 284; Gaps 46;

QY 295 KAATIKAEFVAETPYWVWNPETARRAGOSVSLCCATGKPRDKTFWYHNDLLDPSL 354
 DB 121 KLIVLREDOLEPSGFNIDMGPOLKVERTRTATMLCAASGNPDE-ITWFDLPLVPDPSA 179
 QY 355 YKHESK-----LVLRKIQHQAGEYFCKAOSDAGAVSKVAQLIVTASDETPCNPPEES 408
 DB 180 SNGRIKQRLSGALQIESSEETDOGKECVATNSAGVRYSSANLYVRYR-----RVAPRF 234
 QY 409 YLIRLPHDCFQNA-TNSFYDYGR-CP-VKTCAGQDNGICRDVAVQCCGISKTEEREI 465
 DB 235 SILPMSHIMPGANNITCVAVGSPMPYKMMQGAED-----LTPEDDM 278
 QY 466 QCSGYTLF-TKYAKECSCQRCETRSIVRGVSAADNGEPMRFGHVYKGNRSVMTGY-K 523
 DB 279 PVGRNVLELTVDKDSANHPCVAMSSL--GVIEAV-----AQITVKSLEPK 321
 QY 524 GFTFLHVPDTERLVLTFTVDRLQKFNVTTKVLPFNKKSAAVHEIKMLRKEPTILEAME 583
 DB 322 APGTMTVENTATSTITTWDSGNPDVSVYVIEYKSKSDGPGYQK-----EDIT 371
 QY 584 TNLPLGVEVDEPMALLET-----PSRSFYRQNGEPIYIGVYKASVTFPLDRN 631
 DB 372 TTRYSIG---GISPSSEYEIWSAVNSIGQPPSESVYTRTGEQAPAR-----PPRN 420
 QY 632 ISTATAQDTLNFIDDEDTPLRTYGMFSVDFRGDEVSEPTLNAGKVKVHLDSFOVKME 691
 DB 421 VOARMLSTTKLVQWEE---PVEPNGLIR-GYRYTYME-----PE 457
 QY 692 HISTVKLMSLNDGTGLMEEGDFKFNQRNRKREDRTFLVGNLEIRERLENDLVESRR 751
 DB 458 H-----PVGNMW-----QKHNYDSDLITTVGSL-----LEDET 484
 QY 752 CVYKVRAYRSEFLP-SEQI-----QGVIVSYINL-EPRIGFLSNPRAMG--RFDVYL- 800
 DB 485 YTVRVIAFTFSGDGLSDPTIOVKTQOGVGPOMNLRAEAKSE-TSTITLWSPPROESTIHK 543
 QY 801 -----TGPNGACVPAFCDDQSPDAYSAVVLASLAGEELQAVESSPKFNPNAIGVPOYL 854
 DB 544 YELLREGHGHREV-----GRTFDPTTSYVVEDLKNPTEYAFRLAR--SPQGLAFTPYV 597
 QY 855 NKLIATRTDHEDDR-----VKTKAFQISMAKPRNSAEESGPIYAFENLACEAPPS 908
 DB 598 RORTLOSKSPADPOVKCVSVRSTALIVSWRPPP---ETHNGALVGY----- 642
 QY 909 AAHFRFYOLEGROYDNTVPFNEDD-----PKSWTEDLIAMPKPMERACIYKVK 959
 DB 643 SVRYR-----PLGSEDPKEKVGNGIPPTTQIILEAKETQYRITTVANT 688
 QY 960 IYGP-----LEVNVSRNMGTHRRVYGLKGIIRDVNRSTRDQDPNVSAALEFKCSGMLY 1015
 DB 689 EYGPRESSPVVVKT-----DEDVPSAPPRKVE----- 716
 QY 1016 DDDRVDRTLVKYL-----PGSGRRASVYNMLHEYLVLNHLPLAVNNDTSYTMLAPL 1067
 DB 717 -AELALNATIRVLMRSPAPGRHQIGRGYQVHYVRMEGAEARGPRIKD-----VMLADA 770
 QY 1068 DPLGHVNYGIYTVDDPTAKETALGRCFDGTSDGSSRIKMSNVGVALTFNCVEROVGRO 1127
 DB 771 QEM-----VJNLOPETAISTIVA-AVTMKDGA-----RSKPRVYVYKGAIV--LGRP 815

QY 1128 SAFOYLSTPAOS-----PAAGTVQGRVPSRRQOPASRGQROSGVYASLRF 1175
 DB 816 TL--SVQOTPEGSLARKEPPACTAEDOVULGRLO-----FGNEDSPRLATLEPF 863

RESULT 10
 US-08-800-825A-5
 ; Sequence 5, Application US/0800825A
 ; Patent No. 5866397

; GENERAL INFORMATION:
 ; APPLICANT: RODAN, GIDEON A.
 ; APPLICANT: SCHMIDT, AZRIEL
 ; APPLICANT: RUTLEDGE, SU JANE
 ; TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
 ; TITLE OF INVENTION: TYROSINE PHOSPHATASE
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: J. MARK HAND - MERCK & CO., INC.
 ; STREET: 126 E. LINCOLN AVE., P.O. BOX 2000
 ; CITY: RAHWAY
 ; STATE: NEW JERSEY

; COUNTRY: USA
 ; ZIP: 07065-0900
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/800,825A
 ; FILING DATE: 14-FEB-1997
 ; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
 ; NAME: HAND, J. MARK
 ; REGISTRATION NUMBER: 36,545
 ; REFERENCE/DOCKET NUMBER: 18992DA
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 732-594-3905
 ; TELEFAX: 732-594-4720

; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1911 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-800-825A-5

Query Match 2.3%; Score 144; DB 2; Length 1911;
 Best Local Similarity 19.5%; Pred. No. 0.0019;
 Matches 186; Conservative 122; Mismatches 362; Indels 284; Gaps 46;

QY 295 KAATIKAEFVAETPYWVWNPETARRAGOSVSLCCATGKPRDKTFWYHNDLLDPSL 354
 DB 121 KLIVLREDOLEPSGFNIDMGPOLKVERTRTATMLCAASGNPDE-ITWFDLPLVPDPSA 179
 QY 355 YKHESK-----LVLRKIQHQAGEYFCKAOSDAGAVSKVAQLIVTASDETPCNPPEES 408
 DB 180 SNGRIKQRLSGALQIESSEETDOGKECVATNSAGVRYSSANLYVRYR-----RVAPRF 234
 QY 409 YLIRLPHDCFQNA-TNSFYDYGR-CP-VKTCAGQDNGICRDVAVQCCGISKTEEREI 465
 DB 235 SILPMSHIMPGANNITCVAVGSPMPYKMMQGAED-----LTPEDDM 278
 QY 466 QCSGYTLF-TKYAKECSCQRCETRSIVRGVSAADNGEPMRFGHVYKGNRSVMTGY-K 523
 DB 279 PVGRNVLELTVDKDSANHPCVAMSSL--GVIEAV-----AQITVKSLEPK 321
 QY 524 GFTFLHVPDTERLVLTFTVDRLQKFNVTTKVLPFNKKSAAVHEIKMLRKEPTILEAME 583
 DB 322 APGTMTVENTATSTITTWDSGNPDVSVYVIEYKSKSDGPGYQK-----EDIT 371


```

Db      689 EVGPGESSVYVVRT-----DEDPAPPRKVE----- 716
Oy      1016 DODRVDRTLVYI-----PGSGCRASVNDMLHEYLVLNHLPLAVNNDTSEYTM LAPL 1067
Db      717 -AEALNMTALIRVLMRSRPAQRQHGQIRGYVHYRMEGAARGPRIKD-----VWLADA 770
Oy      1068 DPLGHNYGIVTVDODPRTAKETALGRCEFDGTSDDSSRIKMSNVGVALTFNCVERQVGRQ 1127
Db      771 QEM-----VITNLOPETAYSLIVA-AYTMKGDA-----RSKPKYVVTIKGAV---LGRP 815
Oy      1128 SAFQYLOSTPAQS-----PAAGTVQGRVPSRRQORASRGQSGVVASLRFP 1175
Db      816 TL--SVQOTPEGSLIARWEPAGTAEDQVLYGRLO-----FGREDSPFLATLEFP 863

```

RESULT 12
PCT-US94-10166-5
Sequence 5, Application PC/US9410166

GENERAL INFORMATION:
APPLICANT: RODAN, GIDEON A
APPLICANT: SCHMIDT, AZRIEL
TITLE OF INVENTION: RUTLEDGE, SU JANE
TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: JOHN W. WALLEN III
STREET: P.O. BOX 2000, 126 E. LINCOLN AVE.
CITY: RAHWAY
STATE: NJ
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/10166
FILING DATE: 09-SEPT-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/122,032
FILING DATE: 14-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: WALLEN, JOHN W III
REGISTRATION NUMBER: 35403
REFERENCE/DOCKET NUMBER: 18992
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-594-3905
TELEFAX: 908-594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1911 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-10166-5

Query Match 2.3%; Score 144; DB 5; Length 1911;
Best Local Similarity 19.5%; Pred. No. 0.0019;
Matches 186; Conservative 122; Mismatches 362; Indels 284; Gaps 46;
Oy 295 KAATTAETAEVRYRAETPYMANNPETKARRAGOSVSLCKKATGKRPDPKYVYHNDTLDPDL 354
Db 121 KLVLEDDQLPSGFPINDGSPQIKYVERIRITATMLCAASGNDDPE-ITWKKFLEPVDISA 179
Oy 355 YKHESK-----LVLRKLOOHQAGEYFCKAQSADGAVSKVAQLIVTASDETFPCNPVYES 408

```

Db      180 SNGRIKOLBSGALQIESSEDTQKTECVATNSAGVRYSSPANLIVRYR-----RVAPRF 234
Oy      409 YLIRLPHDFEQNA-TNSFYDYGR-CP-VKTCAGQODNGIRKRAVONCCGISKTEKEI 465
Db      235 SILPMSHEIMPGGNVITFCVAVGSSPMRYVKMAGAD-----LTFEDDM 278
Oy      466 QCSGYTLP-TKVAKEGSCQRCETFSIVGRVSAADNGEPMRFCHVYMGNSRSTGY-K 523
Db      279 PVGRNVLELTDVDSAYAHPCVAMSSL--GVIEAV-----AQITVKSILPK 321
Oy      524 GTFTLHPQDTERLVLFVDRLOKFEVNTTKVLPFNKKSVAHEIKMLRKREPTLEAME 583
Db      322 APGTPMTENTATISITLTMDSGNDDPVSYVYIEYKRSODGPQIR-----EDIT 371
Oy      584 TNIITPLGEVGEDPMARELT-----PSRSFYRONGEPYIGRKAVATFLDPEN 631
Db      372 TTRYSIG--GLSPNSEYELVWSVANSIGGPPSESVVTFTRGQAPAR-----PPRN 420
Oy      632 ISTATAQTDINFTNDEGDFPLRTYGMFSVDREDEYTSPLNAGKVYKHLSTQVKME 691
Db      421 VOARMLSATMTLVQWEE-----PVEPNGLIR-GYRVYITME-----PE 457
Oy      692 HISTVKLMSLNPDGLWEEBDEFEENQRNKRREDRTFLVGNLEIRERLFLNDVPSRR 751
Db      458 H-----PVGNW-----QKHVDDSLFTVGSU-----LEDET 484
Oy      752 CFVKRAYRSEHFLP-SBOI-----OGVVISYINL--EPRTGFSNBRANG--RFDVSI- 800
Db      485 YTVRYLAFTSVGDGLPDLPIQYTKQGVGPQPMNIRARANE--TITLSNPPROBEIIR 543
Oy      801 -----TGPNGACVPAFCDDQSPDAYSAVYLASLAGEELQAVESSPKFNPNAIGVPOYL 854
Db      544 YELLFREGDHGRV-----GRTFDPTTSYVBDLKNPTEVAFRLAAR-SFOGLGAFTPVY 597
Oy      855 NKLINRYRTDHEDEP-----VKTAPOISMAKPRNSAEBSNGPIYAFENLRACEAPPS 908
Db      598 RQRTLOKSPAPPODYKCVSVSTALIVSWRPDP--ETHNGALVGT-----642
Oy      909 AAHFREFQLEGDYVDYNTVPFNEDD-----PMSWTEDYLANPKPMEFRACYIKV 959
Db      643 SVRR-----PLSEDEPEKVENGIPTTQIILLELWTOYRITTVANT 688
Oy      960 IVGP-----LEVNVRSRMGSGTHRRVTGKLYGIRDVYSTDRDROPNVSAACEFKCSGMLY 1015
Db      689 EVGPGESSVYVVRT-----DEDPAPPRKVE-----716
Oy      1016 DODRVDRTLVYI-----PGSGCRASVNDMLHEYLVLNHLPLAVNNDTSEYTM LAPL 1067
Db      717 -AEALNMTALIRVLMRSRPAQRQHGQIRGYVHYRMEGAARGPRIKD-----VWLADA 770
Oy      1068 DPLGHNYGIVTVDODPRTAKETALGRCEFDGTSDDSSRIKMSNVGVALTFNCVERQVGRQ 1127
Db      771 QEM-----VITNLOPETAYSLIVA-AYTMKGDA-----RSKPKYVVTIKGAV---LGRP 815
Oy      1128 SAFQYLOSTPAQS-----PAAGTVQGRVPSRRQORASRGQSGVVASLRFP 1175
Db      816 TL--SVQOTPEGSLIARWEPAGTAEDQVLYGRLO-----FGREDSPFLATLEFP 863

```

RESULT 13
US-08-313-288B-19
Sequence 19, Application US/08313288B
Patent No. 5750502
GENERAL INFORMATION:
APPLICANT: Jessell, Thomas M. and Avihu Klar
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SRONDIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York

COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,288B
FILING DATE: January 5, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40028-A-PC-T-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0526
TELEX:
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1172 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-313-288B-19

Query Match 2.3%; Score 143; DB 1; Length 1172;
Best Local Similarity 20.8%; Pred. No. 0.001;
Matches 105; Conservative 56; Mismatches 189; Indels 154; Gaps 29;

QY 86 CARPLRLAAR-----TDWTPAGSTGVVHSGFRGECFCLNREORPCQ 128
DB 360 CASPFVEGECPCSLHSVDEEGSWPMAEWTCQSVT-----CGSGTQQRGR 406
QY 129 NCSVTVAFILCP-----PGSLRDRTERI-----WSPSPWKSASACGOTGVOTFTRICL 178
DB 407 SCDVTSNTCLPSTQTRACSLSKCDTRIRQDGGMSHSPSSCSYTCG-VGNTRIRICLN 465
QY 179 AEMVSL-----CSEASEEGOHCMGQDC-----TACDLTCPMG--GVNADCD-- 217
DB 466 SPVPOMGKNGKSGSRETKACGAPCPIDGRMSPWSPMSACTVTCAGGIRRETRCNPE 525
QY 218 -----AC-----NC--ODFMLHGAVALP--GGAPASGAALYLLTKPKLLTQTDSD 259
DB 526 PLYGKACVGVDOEROMCNKRSCPDGCLSNPCFFGAGCS-----SPPD 569
QY 260 GRFRIPGLCP-----DGSLIKITKVKFAPYVLTMPKTSLSKAATIKAFVRAETP--YMW 312
DB 570 GSWMC-GCPVPVGLNGHCEDLDECALVPDICSTSKVPKCVNTOPGFHCLPCPPRRYG 628
QY 313 MNPETKARRAGOSVSLCKATGKPRDKYFYVYHNDTLDPSTLYKHESKLVLRKLQHQAG 372
DB 629 NQPRGVGLEAKTEKQVEPE-NPCOKD-----THN-----CHKNAECIYLGHFEDPM-- 675
QY 373 EYFCKASDAGAVSKVAQLIVTASDETPCNPVPESEYILRLPHDFCONATNSFYDV--GR 431
DB 676 -YKCECQTG-----YAGDGLICG--EDSDLDGPNLILVCAATATATYHICIDN 719
QY 432 CPVTCAGQOONGIRCRDAVONCC-----GISTEEREIQCSTYLTPTVAKSCS--- 481
DB 720 CPHLPNSQDEPD--KGLIDACDDDDNDGV--TDEKD--NCQLLFNPROADYDKDEVG 773
QY 482 --CORCTETRSIVRGVSAADNGE 503
DB 774 DRCONCPYVHN--PAQIDTDNNGE 795

RESULT 14
US-08-469-486-52
; Sequence 52, Application us/08469486

Patent No. 5739281
GENERAL INFORMATION:
APPLICANT: Thoesersen, Hans Christian
APPLICANT: Holte, Thor Las
APPLICANT: Etzerodt, Michael
TITLE OF INVENTION: Improved method for the refolding of
TITLE OF INVENTION: proteins
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,486
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/192,060
FILING DATE: February 4, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06363/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 542 8906
TELEFAX: 617 542 5070
TELEX: 200154
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 454 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-486-52

Query Match 2.3%; Score 142.5; DB 1; Length 454;
Best Local Similarity 20.1%; Pred. No. 0.012;
Matches 153; Conservative 73; Mismatches 251; Indels 283; Gaps 39;

QY 132 NYTVAFILCPG-----SLRDRTERIWSFSPWKSASACGOTGVOTFTRICLAEM----- 181
DB 2699 NY---FACPSRCILPMSWTCKEDCEHGEDETHCNKFCSEAOFCQNHRCISKQWLCDG 2755
QY 182 VSLSEASEEGOHCMGQDCADTLCPMGGOVNADDCAGMCDFMLHGAVALPGLGAPASGA 241
DB 2756 SDGCGSDDEAHGCKTCGSSFSFGPTHV-----CVPERWLCDDKCAQDADEBIA 2809
QY 242 AIYLLTKT-----PK-LTQTDSDGRFRIPGLCPDGKSLIKITKVFAPIV 286
DB 2810 AGCLYNSTCDREFCQNRQCIKPFVCDHNRD-----CADGSD----- 2848
QY 287 LTMFKTSLSKAATIKAFVRAETPYVWNPETKARRAGOSVSLCKATGKPRDKYFYVHN 346
DB 2849 -----ESPECEYPTGCPSEFRC--ANGR----- 2869
QY 347 DTLDPSTLYKHESKLVLRKLQHQAGEYFCKAQSAGAVSKVAQLIVTASDETPCNP-- 404
DB 2870 -----CLSSRQWECDEGNDCHDQ-----SDEAKNPHC 2897
QY 405 -VPESEYILRLPHDFCONATNSFYDVGRCPYKT--CAGQOONGIRCRDAV--ONCGISKT 460
DB 2898 TSPE-----HKC--NMSQFLSGSGRCVAELLCNGQDD-----CGDSDBRGCHINEC 2944

Tue Oct 9 12:15:58 2001

us-09-609-383-2.ra1

Page 12

QY 461 EEREIQ-CSGYTLPTKVAKECSQCRCTETRSIVRGVSAADNGEPMRGVHYMGNSRVSM 519
DB 2945 LSRKLSGCSQDCEDLKIGFKRCR-----RPGFRLKDDGFRG----- 2980
QY 520 TGYKGFTHLVPODTERLVLFEVDRLQKFEVNTTKVLPFNKKSAAVHEIKMLRREKPTTL 579
DB 2981 -----ADVDECSITTF-PCSQRCINT-----HGS-----YKCLC----- 3007
QY 580 EAMETNIIPLEGVYGDPMALLETIPSRSEYRONGEPYIGKVASVTFDPRNISTATAQ 639
DB 3008 -----VEGYAP-----RGDPHSCK-----AVTDEEPLI-----FANR 3036
QY 640 TDLNFINDEGDETPPLRTYGM-----FSVDFRDEVT-----SEPLNAGV-----KVHLDSTQVKAP 690
DB 3037 YLLKRLNIDGSNTYLLKQGLNNAVALDFYRDEMITYWDTVTQGSIMIRMHNLGNSVQVL 3096
QY 691 EHISTVKLMSINPDTGL-----WEEGDFKFEENRRKREDRTFLVGNLEIRER 740
DB 3097 HRTGL-----SNPD-GLAVDWVGSLYWCCKGRDTEVSLKNGAY-RTVLVSS-GLREPR 3148
DB 741 LFNLDVPSRRCFYKVAYRSEERFLPSEOIGVIVSYINLEPRTGLSNPRAMGRFDSYI 800
DB 3149 ALVVDVONG-----YLTWTDMDGHSILIGRIGMGSSRSYI-----VDTKI 3188
QY 801 TGPNGACVPAFCDD-QSPDAYSAYV-LASLAGEELQAVES 838
DB 3189 TWPNGLTLDYTERLYMADAREDIIEFASLDGNSRHVYLS 3228
RESULT 15
US-08-469-658-52
Sequence 52, Application US/08469658
Patent No. 5917018
GENERAL INFORMATION:
APPLICANT: Th egeresen, Hans Christian
APPLICANT: Holter, Thor Ias
APPLICANT: Ezerodt, Michael
TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOOLDING OF
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
City: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version
SOFTWARE: #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,658
FILING DATE: June 5, 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/192,060
FILING DATE: February 4, 1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06363/002002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 542 8906
TELEFAX: 617 542 8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 4544 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-658-52
Query Match 2.3%; Score 142.5; DB 2; Length 4544;
Best Local Similarity 20.1%; Pred No. 0.012;
Matches 153; Conservative 73; Mismatches 251; Indels 283; Gaps 39;
QY 132 NYTVRFCLPPG-----SLRDRERIMSPWSPKSCSAACGQTGVOTRRICLAEM----- 181
DB 2699 NY-----FACPSGRCLPMSWTCDEKEDCEHEDETHCKKFCSEAFQCNHRCISKQMLCDG 2755
QY 182 VSLCSEASEGQCHMODCTACDLCTGPMGVNADCDACMCODFMILGAVSLPGAPASGA 241
DB 2756 SDDCGDGSDEAHCCKGTCGPPSSFCPTHV-----CYPERLDCGDKADCADDESIA 2809
QY 242 AYLTLTKT-----PK-LLTQTSDSGRPRIPQLCPDGKSLIKTKVFAPIY 286
DB 2810 AGCLYNSTCDREPMQCNQCIRPKHFVCDHDDRD-----CADGSD----- 2848
QY 287 LTMKTSIKAAITAEFYRAETPYMWNMPETKARBAQSVSLCKKATGRPRPDXYWYHN 346
DB 2849 -----ESPECEYPTGCPSEFRG-----ANGR----- 2869
QY 347 DTLDPSTYKHESKLVLRKLOHQAGEFYCKAQSDAGAVSKVAOLIVTASDETPCNP-- 404
DB 2870 -----CLASSRQWECDEGENDCHDQ-----SDRAPKPNPC 2897
QY 405 -VPESTYLRPLPDGFRONNPNPFYDVGRCPYKT--CAGQDNGIRCDAY-QNCCGISKT 460
DB 2898 TSEP-----HKC--NASSQPLCSSGRCAVALLCNGQDD-----CGDSSDERGCHINEC 2944
QY 461 EEREIQ-CSGYTLPTKVAKECSQCRCTETRSIVRGVSAADNGEPMRGVHYMGNSRVSM 519
DB 2945 LSRKLSGCSQDCEDLKIGFKRCR-----RPGFRLKDDGFRG----- 2980
QY 520 TGYKGFTHLVPODTERLVLFEVDRLQKFEVNTTKVLPFNKKSAAVHEIKMLRREKPTTL 579
DB 2981 -----ADVDECSITTF-PCSQRCINT-----HGS-----YKCLC----- 3007
QY 580 EAMETNIIPLEGVYGDPMALLETIPSRSEYRONGEPYIGKVASVTFDPRNISTATAQ 639
DB 3008 -----VEGYAP-----RGDPHSCK-----AVTDEEPLI-----FANR 3036
QY 640 TDLNFINDEGDETPPLRTYGM-----FSVDFRDEVT-----SEPLNAGV-----KVHLDSTQVKAP 690
DB 3037 YLLKRLNIDGSNTYLLKQGLNNAVALDFYRDEMITYWDTVTQGSIMIRMHNLGNSVQVL 3096
QY 691 EHISTVKLMSINPDTGL-----WEEGDFKFEENRRKREDRTFLVGNLEIRER 740
DB 3097 HRTGL-----SNPD-GLAVDWVGSLYWCCKGRDTEVSLKNGAY-RTVLVSS-GLREPR 3148
DB 741 LFNLDVPSRRCFYKVAYRSEERFLPSEOIGVIVSYINLEPRTGLSNPRAMGRFDSYI 800
DB 3149 ALVVDVONG-----YLTWTDMDGHSILIGRIGMGSSRSYI-----VDTKI 3188
DB 3189 TWPNGLTLDYTERLYMADAREDIIEFASLDGNSRHVYLS 3228
Search completed: October 9, 2001, 11:50:10
Job time: 265 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2001, 11:46:55 ; Search time 24.75 Seconds
(without alignments)
3644.067 Million cell updates/sec

Title: US-09-609-383-2

Perfect score: 6317
Sequence: 1 MVGTRAWVPSFLVLEVTSVL.....QSGVVASLRFPRAQGPLIN 1184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6317	100.0	1184	2 T09484	cartilage intermed
2	169	2.7	1170	2 A40558	thrombospondin 1 p
3	161	2.5	1694	2 S50065	stathomedhesin - mou
4	157	2.5	1170	1 TSH0P1	thrombospondin 1 p
5	155.5	2.5	7962	2 I38346	elastic titin - hu
6	152	2.4	6658	2 T13931	projectin - fruit
7	148.5	2.4	1125	1 JH0771	protein-tyrosine k
8	147.5	2.3	788	2 T25061	hypothetical prote
9	147	2.3	1444	2 A32579	angiogenesis inh
10	146	2.3	1239	1 A32579	neuroglian - fruit
11	145	2.3	1906	1 S68235	myosin-11ght-chain
12	144.5	2.3	1176	2 JN0583	myosin-11ght-chain
13	143.5	2.3	1034	2 JCS598	mucin - rat
14	143	2.3	1172	1 TSH0P2	thrombospondin 2 p
15	143	2.3	1360	2 T33922	hypothetical prote
16	142.5	2.3	4544	1 S02392	alpha-2-macroglobu
17	142	2.2	1074	1 JCS928	semaphorin F precu
18	142	2.2	3570	2 T45025	mucin MUC5B, trach
19	141.5	2.2	477	2 S53362	mucin 5AC (clone J
20	141	2.2	1236	2 T03096	CD0 protein - rat
21	141	2.2	6642	2 T29757	protein UNC-89 - C
22	140.5	2.2	3375	2 T19821	hypothetical prote
23	140	2.2	1122	2 I54237	protein-tyrosine k
24	139.5	2.2	1056	2 A53767	mucin MUC5B, trach
25	139.5	2.2	4391	2 A38096	pelican precursor
26	139	2.2	984	2 T00326	hypothetical prote
27	139	2.2	1522	2 T00028	brain-specific ang
28	139	2.2	2222	2 T13924	sdh protein - fru
29	139	2.2	4162	2 T42633	connectin/titin -

30	138.5	2.2	1123	1 JN0712	protein-tyrosine k
31	138	2.2	1599	2 T16210	hypothetical prote
32	136	2.2	2167	2 T34395	hypothetical prote
33	136	2.2	2907	2 A57278	fibroblastin-2 precu
34	135	2.1	1147	2 A59307	myosin-11ght-chain
35	135	2.1	2918	2 A54105	fibroblastin-2 precu
36	134.5	2.1	1450	2 A44027	165k myofibrillar
37	134	2.1	1651	2 T14160	transmembrane rece
38	133.5	2.1	1059	2 T22545	hypothetical prote
39	133	2.1	1894	2 C54689	hypothetical prote
40	132.5	2.1	610	2 T16761	protein-tyrosine-p
41	132.5	2.1	860	2 T16892	hypothetical prote
42	132.5	2.1	4545	1 S25111	hypothetical prote
43	132	2.1	1178	1 A39804	alpha-2-macroglobu
44	131.5	2.1	1742	2 S24600	thrombospondin pre
45	131.5	2.1	3707	2 S18252	projectin - fruit heparan sulfate pr

ALIGNMENTS

RESULT 1
T09484
cartilage intermediate layer protein precursor - human
C:Species: Homo sapiens (man)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T09484
R:Lorenzo, P.; Neame, P.; Sommarin, Y.; Heinegard, D.
J. Biol. Chem. 273, 23469-23475, 1998
A:Title: Cloning and deduced amino acid sequence of a novel cartilage protein (CILP)
A:Reference number: 216689; MUID:98389785
A:Accession: T09484
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1184 <LOR>
A:Cross-references: EMBL:AF035408; NID:g3513502; PIDN:AAC33838.1; PID:g3513503
A:Experimental source: tissue type articular cartilage
C:Genetics:
A:Note: CILP
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1184/Product: cartilage intermediate layer protein #status predicted <MAT>
Query Match 100.0%; Score 6317; DB 2; Length 1184;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MVGTRAWVPSFLVLEVTSVLQRTMLTOSVRVQGGKKNPSIFAKPADTLESFGEMTTLF 60
|||||
DB 1 MVGTRAWVPSFLVLEVTSVLQRTMLTOSVRVQGGKKNPSIFAKPADTLESFGEMTTLF 60
OY 61 NIDYGGKGDYERLDAIRFYGGDRVCARPLRLAARTTMTFAGSTGVVHGSFRCGFMC 120
|||||
DB 61 NIDYGGKGDYERLDAIRFYGGDRVCARPLRLAARTTMTFAGSTGVVHGSFRCGFMC 120
OY 121 NREQPRGNCSTYTRFLCPGSLRPRDTERITSPSPSKSCGAAGGQGVQVQRTTICLAE 180
|||||
DB 121 NREQPRGNCSTYTRFLCPGSLRPRDTERITSPSPSKSCGAAGGQGVQVQRTTICLAE 180
OY 121 NREQPRGNCSTYTRFLCPGSLRPRDTERITSPSPSKSCGAAGGQGVQVQRTTICLAE 180
|||||
DB 121 NREQPRGNCSTYTRFLCPGSLRPRDTERITSPSPSKSCGAAGGQGVQVQRTTICLAE 180
OY 181 MVSLSSEASEEGOHCMGDCITACDLTCFPMGVNACDCMCMODFMLHGAVSLPGAPASG 240
|||||
DB 181 MVSLSSEASEEGOHCMGDCITACDLTCFPMGVNACDCMCMODFMLHGAVSLPGAPASG 240
OY 241 AAIYLLFTKPTKLLTQTDSDGRRIRPGICPDGKSLITKRVKAPIVLTPMKTSLKAATIK 300
|||||
DB 241 AAIYLLFTKPTKLLTQTDSDGRRIRPGICPDGKSLITKRVKAPIVLTPMKTSLKAATIK 300
OY 301 AEFVAETPYVMNPETAKRARGOSVSLCCRTGKPRDQKRYVYHNDLLDPSLTKKHESK 360
|||||
DB 301 AEFVAETPYVMNPETAKRARGOSVSLCCRTGKPRDQKRYVYHNDLLDPSLTKKHESK 360
OY 361 LVLRKLOOHQAGEYFCCKAQSDAGAVSKVAQDLIVASDETPCNPVPESTLIRLPDCEFN 420
|||||

Db 361 LVKRLQHQAGEFCAQSAQAVKSVLAQLIVTASDEPCNPVPSYLLIRLPHDCFQN 420
 QY 421 ATNSFYDVGRCPYKTAGGQDNGIRCDAYONCGISKTREERIGSGYTLPTKVAEC 480
 Db 421 ATNSFYDVGRCPYKTAGGQDNGIRCDAYONCGISKTREERIGSGYTLPTKVAEC 480
 QY 481 SCQCTETRSIVRGRVSAADNGEPMRGHYMGNSRVSMIGYKGTFTLHVPODTERLVLT 540
 Db 481 SCQCTETRSIVRGRVSAADNGEPMRGHYMGNSRVSMIGYKGTFTLHVPODTERLVLT 540
 QY 541 FVDRLOKFVNTTKVLPFNKKGSFAVHEIKMLRKREPTLEAMENITPLGVVEDPMAE 600
 Db 541 FVDRLOKFVNTTKVLPFNKKGSFAVHEIKMLRKREPTLEAMENITPLGVVEDPMAE 600
 QY 601 LEIDRSFYRONGEPYIGKVAASYTLDPDRNISTATAOTDLNFINDEGDTFPLRTYGMF 660
 Db 601 LEIDRSFYRONGEPYIGKVAASYTLDPDRNISTATAOTDLNFINDEGDTFPLRTYGMF 660
 QY 661 SVDRDVTSPPLNAGVKVYVLDSTQYKMPHISTVKLMSLNPTGLWEEGDKFENOR 720
 Db 661 SVDRDVTSPPLNAGVKVYVLDSTQYKMPHISTVKLMSLNPTGLWEEGDKFENOR 720
 QY 721 RNKREDRTFLVGNLEIERERLFINLDPESRRCFYKVAYSERFLPSEIOGVVITVINL 780
 Db 721 RNKREDRTFLVGNLEIERERLFINLDPESRRCFYKVAYSERFLPSEIOGVVITVINL 780
 QY 781 EPRGTGFLSNPRAGRPDSVITGPGACVPAFCDDQSDPAYSAYVLASLAGEELQAVESSP 840
 Db 781 EPRGTGFLSNPRAGRPDSVITGPGACVPAFCDDQSDPAYSAYVLASLAGEELQAVESSP 840
 QY 841 KFNNAIGVPOPYLNKNTYRTHEDPRVKTAFOISMARPRPASAESNPIYAFENLR 900
 Db 841 KFNNAIGVPOPYLNKNTYRTHEDPRVKTAFOISMARPRPASAESNPIYAFENLR 900
 QY 901 ACEBAPSAAHFREFYQIEGDRYDNTVPFNEEDPMSTEDYLAMPMPMEFRACYITVKI 960
 Db 901 ACEBAPSAAHFREFYQIEGDRYDNTVPFNEEDPMSTEDYLAMPMPMEFRACYITVKI 960
 QY 961 VGPLLEVAVNRNMGTHRRVYKLYGIRDYRSTFRDQDPVNSAACLEFFKCSGMLYDDRY 1020
 Db 961 VGPLLEVAVNRNMGTHRRVYKLYGIRDYRSTFRDQDPVNSAACLEFFKCSGMLYDDRY 1020
 QY 1021 DRTLVKTIPOGSCRASVYNMLHEYLIVNHLPLAVNNDTSEYTMALPDLPGHNGITVT 1080
 Db 1021 DRTLVKTIPOGSCRASVYNMLHEYLIVNHLPLAVNNDTSEYTMALPDLPGHNGITVT 1080
 QY 1081 DDPRTAKETALGRCFEDTSDGSSRIKSNVGVALTFCNCEYRGVGRSARFOYLQSTPAQS 1140
 Db 1081 DDPRTAKETALGRCFEDTSDGSSRIKSNVGVALTFCNCEYRGVGRSARFOYLQSTPAQS 1140
 QY 1141 PAAGTVQGRVPSPRQOASRGQROSGVASLPRPVAQOPLIN 1184
 Db 1141 PAAGTVQGRVPSPRQOASRGQROSGVASLPRPVAQOPLIN 1184

RESULT 2
 A:Accession: A40558
 C:Species: Mus musculus (house mouse)
 C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #extl_change 20-Aug-1999
 C:Accession: A40558; A37905; B42587; S68787
 R:Lawler, J.; Duquette, M.; Petro, P.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.
 Genomics 11, 587-600, 1991
 A:Title: Characterization of the murine thrombospondin gene.
 A:Reference number: A40558; M01D:92128941
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1170 <LAN>
 A:Cross-references: GB:M62449; GB:M62450; GB:M62451; GB:M62452; GB:M62453; GB:M62454; GB:M62455; GB:M62466; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:9511867; PIDN:AAA5061
 J:Bornstein, P.; Alt, D.; Devaiah, S.; Framson, P.; Li, P.
 J. Biol. Chem. 265, 16691-16698, 1990

A:Title: Characterization of the mouse thrombospondin gene and evaluation of the role
 A:Reference number: A37905; M01D:90375546
 A:Accession: A37905
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-490 <BOR>
 A:Cross-references: GB:J05605; GB:J05606; NID:9201991; PIDN:AAA40431.1; PID:9554390
 R:Liherly, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seidlin, M.F.; Dixit, V.M.
 J. Biol. Chem. 267, 3274-3281, 1992
 A:Title: Characterization of mouse thrombospondin 2 sequence and expression during ce
 A:Reference number: A42587; M01D:92147683
 A:Accession: B42587
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1152, 'P', 1154-1170 <LAN>
 A:Cross-references: GB:M87276
 A:Note: sequence extracted from NCBI backbone (NCBI:81501)
 R:Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D.F.
 FEBS Lett. 387, 36-41, 1996
 A:Title: Expression and initial characterization of recombinant mouse thrombospondin
 A:Reference number: S68787; M01D:96234006
 A:Accession: S68787
 A:Molecule type: protein
 A:Residues: 19-26, 'X', 28-37 <CHE>
 C:Complex: homotrimer, disulfide linked
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;
 C:Keywords: calcium binding; glycoprotein; homotrimer
 F:1-18/Domain: signal sequence #status predicted <Sts>
 F:19-1170/Product: thrombospondin 1 #status predicted <NAT>
 F:317-375/Domain: von Willebrand factor type C repeat homology <WMC>
 F:378-429/Domain: thrombospondin type 1 repeat homology <THR1>
 F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>
 F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
 F:551-586/Domain: EGF homology <EGF>
 F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.7%; Score 169; DB 2; Length 1170;
 Best local similarity 20.6%; Pred. No. 0.0028;
 Matches 113; Conservative 38; Mismatches 151; Indels 246; Gaps 28;

QY 10 SFLVLEVTSLVGLQMTLQSVRRVQEK-----KNPSFAKPAOTLSPGCGWTT----- 58
 Db 278 SSMVLELKGRLTIVTLQDSIRKTEENRELVSRLRPPLCFHNGVQYKNEEMTVDSCT 337
 QY 59 -----LFNIDYPGKGADYERLDALREFYGGDVCARPLLEARTTD 98
 Db 338 ECHQNSVITICKRVSCPIMCSNATYPDG-----CPRCWFSDSADG 381
 QY 99 WTPAGSTGVVHGSFRRGEW-----CLNREQRPGNCSNYTVRFCLPGSLRDRTERI- 151
 Db 382 WSPWSE-----WTSQATGNGIQGRSCDILNR--CEGSSVQTRCHIQ 426
 QY 152 -----WSPWSPWMSCSAAGOTGVQTRTCLAEWML-----CSEASEEOHOM 196
 Db 427 ECKRREKQDGMWSHSPWSSCVTCGD-GVITRIRLCSNPSPOGNKPKCEGEARETRKAK 485
 QY 197 GQDC-----TACDLTC-----PMGOVNA-DC-----DACMCD 223
 Db 486 KDACPINGMGWPMSPMDICSVTCGGVQRRSRCLNPTQFGKDCVGVTEQVYCNKRD 545
 QY 224 FMHAGVSLPGCAPASGAATLYLTTPKLLTQDSGRFRIRPGLCDGKSTIKTKVKPA 283
 Db 546 CPIDGCLSNPCFAGAK-----CTSTP-----DGSWKC-GACPPGY- 580
 QY 284 PVLTPKTSLSKATIKAEVRAETPYMMNPEPTKARRAGQSYSLCKKATG--KRPDKY 341
 Db 581 -----GNGIQ--CKDYDECKEVPDAG 599
 QY 342 FWHYNDTLDPSSLYKHESKLVLRKLOHQAGEYFCKAQSDAG----- 383
 Db 600 F-NHN-----GERHCK-NTPDPGYNCLPCCPRRTGSGSPQR 632

[illegible]

RESULT 4

TSMDP1

thrombospondin 1 precursor - human

C:Species: Homo sapiens (man)

C:Date: 23-Aug-1987 #sequence (man)

C:Revision: 03-Aug-1995 #text change 17-Nov-2000

A:Accession: A26155; A34274; A30140; A25812; A05172; A42927

R:Lawler, J.; Hynes, R.O.

J. Cell Biol. 103, 1635-1648, 1986

A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multiple

A:Reference number: A26155; MUID:87057617

A:Accession: A26155

A:Molecule type: mRNA

A:Residues: 1-1170 <LMB>

A:Cross-references: GB:X04665; NID:937137; PIDN:CAA82370.1; PID:937138

A:Note: parts of this sequence, including the amino end of the mature protein, were d

J. Laberth, C.D.; Gierman, T.M.; Dixit, V.M.

J. Biol. Chem. 264, 11222-11227, 1989

A:Title: Characterization of the promoter region of the human thrombospondin gene. DN

A:Reference number: A34274; MUID:89291870

A:Accession: A34274

A:Molecule type: DNA

A:Residues: 1-166 <LAR>

A:Cross-references: GB:J04835

R. Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwel

J. Cell Biol. 108, 729-736, 1989

A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in

A:Reference number: A30140; MUID:89139590

A:Accession: A30140

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-522, 'A', '524-1170 <HEN>

A:Cross-references: EMBL:X14787; NID:937464; PIDN:CAA32889.1; PID:937465

A:Note: parts of this sequence, including the amino end of the mature protein, were d

R. Kobayashi, S.; Eden-McCutchan, F.; Framson, P.; Bornstein, P.

Biochemistry 25, 8418-8425, 1986

A:Title: Partial amino acid sequence of human thrombospondin as determined by analysi

A:Reference number: A25812; MUID:87157592

A:Accession: A25812

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-397 <KOB>

A:Cross-references: GB:M2631; NID:9538353; PIDN:AAA63741.1; PID:9538354

R. Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986

A:Reference number: A05172; MUID:86287276

A:Accession: A05172

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-374, 'RC' <DIX>

A:Cross-references: GB:M1326; NID:9340005; PIDN:AAA61237.1; PID:9553801

A:Note: parts of this sequence, including the amino end of the mature protein, were d

R. Sun, X.; Skorstengaard, K.; Mosher, D.F.

J. Cell Biol. 118, 693-701, 1992

A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin

A:Reference number: A42927; MUID:92348511

A:Accession: A42927

A:Molecule type: Protein

A:Residues: 987-1003 <SUN>

A:Note: Cys-992 is shown to have a free sulfhydryl

A:Accession: A26155

A:Molecule type: mRNA

A:Residues: 1-1170 <LMB>

A:Cross-references: GB:X04665; NID:937137; PIDN:CAA82370.1; PID:937138

A:Note: parts of this sequence, including the amino end of the mature protein, were d

J. Laberth, C.D.; Gierman, T.M.; Dixit, V.M.

J. Biol. Chem. 264, 11222-11227, 1989

A:Title: Characterization of the promoter region of the human thrombospondin gene. DN

A:Reference number: A34274; MUID:89291870

A:Accession: A34274

A:Molecule type: DNA

A:Residues: 1-166 <LAR>

A:Cross-references: GB:J04835

R. Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwel

J. Cell Biol. 108, 729-736, 1989

A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in

A:Reference number: A30140; MUID:89139590

A:Accession: A30140

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-522, 'A', '524-1170 <HEN>

A:Cross-references: EMBL:X14787; NID:937464; PIDN:CAA32889.1; PID:937465

A:Note: parts of this sequence, including the amino end of the mature protein, were d

R. Kobayashi, S.; Eden-McCutchan, F.; Framson, P.; Bornstein, P.

Biochemistry 25, 8418-8425, 1986

A:Title: Partial amino acid sequence of human thrombospondin as determined by analysi

A:Reference number: A25812; MUID:87157592

A:Accession: A25812

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-397 <KOB>

A:Cross-references: GB:M2631; NID:9538353; PIDN:AAA63741.1; PID:9538354

R. Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986

A:Reference number: A05172; MUID:86287276

A:Accession: A05172

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-374, 'RC' <DIX>

A:Cross-references: GB:M1326; NID:9340005; PIDN:AAA61237.1; PID:9553801

A:Note: parts of this sequence, including the amino end of the mature protein, were d

R. Sun, X.; Skorstengaard, K.; Mosher, D.F.

J. Cell Biol. 118, 693-701, 1992

A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin

A:Reference number: A42927; MUID:92348511

A:Accession: A42927

A:Molecule type: Protein

A:Residues: 987-1003 <SUN>

A:Note: Cys-992 is shown to have a free sulfhydryl

A:Accession: A26155

A:Molecule type: mRNA

A:Residues: 1-1170 <LMB>

A:Cross-references: GB:X04665; NID:937137; PIDN:CAA82370.1; PID:937138

A:Note: parts of this sequence, including the amino end of the mature protein, were d

J. Laberth, C.D.; Gierman, T.M.; Dixit, V.M.

J. Biol. Chem. 264, 11222-11227, 1989

A:Title: Characterization of the promoter region of the human thrombospondin gene. DN

A:Reference number: A34274; MUID:89291870

A:Accession: A34274

A:Molecule type: DNA

A:Residues: 1-166 <LAR>

A:Cross-references: GB:J04835

R. Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwel

J. Cell Biol. 108, 729-736, 1989

A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in

A:Reference number: A30140; MUID:89139590

A:Accession: A30140

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-522, 'A', '524-1170 <HEN>

A:Cross-references: EMBL:X14787; NID:937464; PIDN:CAA32889.1; PID:937465

A:Note: parts of this sequence, including the amino end of the mature protein, were d

R. Kobayashi, S.; Eden-McCutchan, F.; Framson, P.; Bornstein, P.

Biochemistry 25, 8418-8425, 1986

A:Title: Partial amino acid sequence of human thrombospondin as determined by analysi

A:Reference number: A25812; MUID:87157592

A:Accession: A25812

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-397 <KOB>

A:Cross-references: GB:M2631; NID:9538353; PIDN:AAA63741.1; PID:9538354

R. Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986

A:Reference number: A05172; MUID:86287276

A:Accession: A05172

A:Molecule type: mRNA

A:Residues: 1-83, 'A', '85-374, 'RC' <DIX>

A:Cross-references: GB:M1326; NID:9340005; PIDN:AAA61237.1; PID:9553801

A:Note: parts of this sequence, including the amino end of the mature protein, were d

R. Sun, X.; Skorstengaard, K.; Mosher, D.F.

J. Cell Biol. 118, 693-701, 1992

A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin

A:Reference number: A42927; MUID:92348511

A:Accession: A42927

A:Molecule type: Protein

A:Residues: 987-1003 <SUN>

A:Note: Cys-992 is shown to have a free sulfhydryl

A:Accession: A26155

A:Molecule type: mRNA

A:Residues: 1-1170 <LMB>

A:Cross-references: GB:X04665; NID:937137; PIDN:CAA82370.1; PID:937138

A:Note: parts of this sequence, including the amino end of the mature protein, were d

J. Laberth, C.D.; Gierman, T.M.; Dixit, V.M.

J. Biol. Chem. 264, 11222-11227, 1989

A:Title: Characterization of the promoter region of the human thrombospondin gene. DN

A:Reference number: A34274; MUID:89291870

A:Accession: A34274

A:Molecule type: DNA

A:Residues: 1-166 <LAR>

A:Cross-references: GB:J04835

R. Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwel

J

F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>
 F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
 F:551-586/Domain: EGF homology <EGF1>
 F:650-689/Domain: EGF homology <EGF2>
 F:926-928/Region: cell attachment (R-G-D) motif
 F:171-232/Disulfide bonds: #status Predicted
 F:248-360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:270,274/Disulfide bonds: interchain #status predicted
 F:610/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
 F:1051/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 2.5%; Score 157; DB 1; Length 1170;
 Best Local Similarity 20.1%; Pred. No. 0.019;

Matches 111; Conservative 37; Mismatches 150; Indels 254; Gaps 27;

```

QY 10 SFVLVETVSLGKQMTLTQSVRYQPKKPSIFAKPADLLESP-----GEW 56
DB 278 SSMVLELRLGLRTVTLTQDSIRKYTEENK-----ELANELRRPPLCYHNGVOYRNNEEW 331
QY 57 TT-----LEFIDYFGKGDYERLDAIRFYGDVRCARPLRL 92
DB 332 TVDSCEHCQNSVTICKKVCSPIMCSNATVDEG-----CCPRCWP 375
QY 93 EARTDWTPTAGSTGVVHGSPPREGFW-----CLNREQRQNCNNTVRLCPPGSLRR 146
DB 376 DSADDGKSPSE-----WTSCSTSCGNGIQGRGSDSLNNR--CEGSSVQT 420
QY 147 DTERI-----WSPWSPWKSACACGQTVOTRTRICLAENVSL-----CSASE 130
DB 421 RTCHIEBCDRKFKODGWSHSPWSSCSVTGCD-GVITRIRLCNSPSPOMNGKRCGEAR 479
QY 191 EGOHCMGDC-----TACDLTC-----PMGOVNA-DC-----D 217
DB 480 ETACKKCKADCPINGMGPMSPMDICSTYCGGQVOKSRCLNNPTPPGGKDCYDVTENQ 539
QY 218 ACMDQDFMLHGAVALSPGADASGAAYLLTKTKRLLTQDSDGRFRIRPGICPDGKSLTKI 277
DB 540 ICKKQDCPDIGCLSNCFACVYK-----CTSP-----DSMKC-GACPPGYS--- 580
QY 278 TKKFAPIVLTMTKSLKATIKAEVRAETPYVMWNPETKARRAGSVSLCKKATGKPR 337
DB 581 -----GNGIQCTDVDECKEY 595
QY 338 PDKFYWHNDTLDPRLSKHESKLVRLKLOHQAGEYFCKAQSADG----- 383
DB 596 PDACF-NHN-----GEHRCE-NTDPGNCILPCPRPRTGSQ 628
QY 384 ----AVKSKYAQLIVTASDETPCNVPYESTYLRLPHDCFONATNSF---YD-VGRCPVK 435
DB 629 PFGQVEHATANKQVC---KPRNPCTDG---THDCNKNACNYLGHYSDEMYRCECK 679
QY 436 TCAGQDNGIRIC 447
DB 680 --PGYAGNGILIC 689

```

RESULT 5

138346 elastic titin - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000

C:Accession: I38346

R:Label: S.; Kolmerer, B. Science 270, 293-296, 1995

A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.

A:Reference number: A57430; M0ID:96026330

A:Accession: I38346

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues:1-17962 <RES>

A:Cross-references: EMBL:X90569; NID:q1017426; PIDD:CAA62189.1; PID:q1017427

C:Genetics:

A:Gene: GDB:TTN
 A:Cross-references: GDB:127867; OMIM:188840
 A:Map position: 2q31-2q31

Query Match 2.5%; Score 155.5; DB 2; Length 7962;

Best Local Similarity 17.5%; Pred. No. 0.4;

Matches 197; Conservative 139; Mismatches 353; Indels 437; Gaps 51;

```

QY 17 TSVLGRQMTLTQSV-----RRVQPKKKNPSIFAKPADLLESGEWTTL-----FN 61
DB 2529 SSDMGNTVCVAAVAVAGSDECAVLTQDEPPSFVKEPEPELVLGKNVTFVIRGTPPEFK 2588
QY 62 IDYFGKGDYERLDAIRFYGDVRCARPLLEARTDWTMPAGSTGVVHGSPPREGWCIN 121
DB 2589 VNNFRGARELVKGDRCNITFEDTVA---ELELFNIDISQSGEYTCVSNNAQAS-CTT 2643
QY 122 R-----EQRPQNC--SNVTRFLCPPLSLRDRTERIWSPPWS-PWSK- 160
DB 2644 RLFVKEPAFLKRLSDHSVSPGKSIILESTYT-----GTL-----PISVTWKKD 2687
QY 161 ----CSACGQTVOTRTRICLAENVSLCSSEASBSQHCMDCTACDILTCPMGOVNA 215
DB 2688 GFNITTSSEK--NIYTEKTCILEILN--STRKDAQY-----SCEIENBAGR--- 2731
QY 216 CDACMGDFMLHGAVALSPGAPASGAAYLLTKPKLLTQDSDGRFRIRPGICPDGKSLI 275
DB 2732 -DYC-----GALVSTL----- 2741
QY 276 KITKVKFAPIVLTMTKSLKATIKAEVRAETPYVMWNPETKARRAGSVSLCKKATGK 335
DB 2742 -----EPVYFTELEPLAAVAGSVSLQCCVAGT 2770
QY 336 PRDQKFWYHNDTLDP-----LKHKESKLVRLKLOHQAGEYFCKAQSADGAVKSV 389
DB 2771 PE-ITVSWYKGDYKRLRTPERYTFYTNNAATLVENKVININDSEYTKAENSIGTASST 2829
QY 390 AOLI-----VTSDETPCNVP-----ESTYLIRLPHDCFO- 419
DB 2830 VFRIQERQLPSPARQLKDIQTVGLPVTLTCRLNSAPLQVWYDGVLLR-DHENLOT 2888
QY 420 -----NATN-----SFYYDVGRCPVKTC 437
DB 2889 SFVDNVAATLKILOTDLSSHSGVSCSASNPLGTASSARLARARPKKSPFDIKPVSDVY 2948
QY 438 AGQDNGICRDAVQMCQGS-KTEEREIQCSG-YILP-----TKVAKEGSCQ-R 484
DB 2949 AGESAD-FECHVTGAOPMRITWSKDKKEIRPGGNVYITCVGNTPHLRILKVCCKGSGQYT 3007
QY 485 CTETRSIVRGVSA-----ADGEMPRFGHYVMGNSRVSMGTGKGT 525
DB 3008 CQATNDVGMKMSAQLSVKEPRFVKKLKASKYAKAGESQLECKISGSEPIKVSFRND 3067
QY 526 FTLLHPQDTERLVLTIVDRLOKQVNTTKVLPNK----- 559
DB 3068 SELHESMKYV-----MSFINSVALITINESAEDSGDYICEAHNGVDASCSTAL 3117
QY 560 --KGSVFEHEIKMLRKEPTITLAME-TNIIPLGEVVGEDPMALFIPRSFYRQGEY 616
DB 3118 TVKAPPVF-----TQKSPYALGSDVILLOCELSGTPPEVAVWVKDKOYR-NSKRF 3169
QY 617 IGRKASVYFLDPRNISTATAQTDLNFINDGE-DTF-----PLRTYGMFS----- 661
DB 3170 KIRSKHEDTNLHLINLEASDVGEYHCKANNVSGSDSCSVKFKRPRVVKLSOTSTLI 3229
QY 662 ---VDRDEYTS-EPLNA-----GKVYVLDSTQVKKMPEHISTYKLS----- 700
DB 3230 GDVAVELRAIVIEGQPTISVWMLKDRGEVIRESENTRISFIDNIATLQIGSPASNSGKYIC 3289
QY 701 -LNPDRGLMEBEDDFEFENQRR--NKREDRTFLVGN----- 733
DB 3290 QIKNDAGMRCSAVLTVLEPARITTEKPEPTVYTTGPNPALCEVVTGTPELSAKWFRDGE 3349

```

```

OY 734 -----LEIRERLRLNDVPE-----RRCFVKVRAKSERELPS- 767
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 3350 LSADSKHHITFINKVASLKIPCAEKSMDKGLVSEFKNSGNCJVSYHV--SDRLIVPS 3407
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 768 -----EQLOGVVISYINLEPRGFLSNBRAMGRF---DSVITGPGACVPAFCD----- 813
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 3408 FIRKLKDNALIGASVLEECRSG--SAPISVGMFODGNEIVSGPK--COSSFENVCITLN 3464
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 814 -----DQSPDAYSAYVLASLAGE-----LQAVESPKE--NPALICV 849
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 3465 LSLPEPDTGITTCVAAVAVAGSDSCSAVLTVQEPSPFQPTDPSVEV 3510
| : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 6
T13931
projectin - fruit fly (Drosophila melanogaster) (fragment)
C:Species: Drosophila melanogaster
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T13931
R:Daley, J.; Southgate, R.; Ayme-Southgate, A.
J. Mol. Biol. 279, 201-210, 1998
A:Title: Structure of the Drosophila projectin protein: isoforms and implication for pro
A:Reference number: Z17815; MUID:98300339
A:Accession: T13931
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Cross-references: EMBL:AF047475; NID:g3337430; PID:g3337431; PIDN:AAC27550.1
A:Residues: 1-6658 <DAL>
A:Molecule type: DNA
A:Gene: projectin
A:Cross-references: FlyBase:FBgn0005666
A:Map position: 4
A:Note: intron positions not resolved (incomplete sequence)
C:Keywords: muscle

```

```

Query Match 2.4%; Score 152; DB 2; Length 6658;
Best Local Similarity 19.6%; Pred. No. 0.54;
Matches 162; Conservative 100; Mismatches 259; Indels 306; Gaps 40;

```

```

OY 234 GAAPASGAAT-----YLLTKPKLLTQTD-----SDGRFRIPGLCPDGK---- 272
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1742 GGASITGYIVRKDPNTGKWKOKALETSTPDCARVNDLIAGNKYQFRIMAVNKAAGSKSPS 1801
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 273 ---SILKIKVFAPIVLTMPK---TSLKATIKAEFAETPYWMANPETKARRAGOSVS 327
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1802 EPSDOMAKDRAP-----PKIDRTNIDITSK-----AGQHTR 1835
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 328 LCCATGKPRPDKYFWYHNDTL--DPSLY-----KHESKLYLRKIQOHAGEYRCQKAS 380
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1836 FDIKVSGER-PATKVMILNKARLENDSDNYNIDMESYRKLTVPISTKRHSKYLTKAEN 1894
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 381 DAGAVKSVQALIVTASDETPCNVPESYLRLPHDCFQMATNSEFYDVGK--CPVKTC 438
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1895 ESGRDEASFEVILVD-----KPGPPGGLPLKVT-----DVHKRGCKLKNWA 1934
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 439 GQGDNGICRQAVONCCISKTEEBEIQSGTTLPTKAKECSQ--KCTETSTIYGRKVS 497
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1935 PLDDGGLPIDHYI-----IEKMDVE--SGRMPLPSGRKESFAELNNLEPSHEKFFRL 1985
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 498 AAD---NEPMFEGHYVGNRSFVSWTGYKGFETLHVPODTERLVLFVDRLOKFPNTTKV 554
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1966 AVNTGSESEPLTGEQSVLAKNPFDEPGKPGT-----PE-----AYDMQKQHDVLWR 2032
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 555 LPEKKGSAVHEIKMLRKKEPTT--LEAMETNIIPGE-----VVGEDPMALLET- 603
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2033 PPINDGSGPITGYV--VEKREKGTAKIKGEITIPICLGECKATVPRLNCEYEFRVK 2090
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 604 -----PSSSFYRQNGEPYIGKV-KASYVFLDP--ENISATATAQDILN--INDG 650
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2091 AINAAQPEPSDA-----SKPIITPKRIAPITLDPITKNIKRYNFKSGEPIFLDINISGE 2145
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 651 TEPRLT-----YGMFSVDR 665
| : : : : : | : : : : : | : : : : : | : : : : : |

```

```

Db 2146 PAPDVTWNNKNSVQTTSPSHLENLPYNTKIINNPERKDTGLYKISAHNFYGGQOVREQ 2205
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 666 DEVTSEPLNAG-----KQVYHLDSTOVKMP-----EHISTVKLMSLNDPTGLMBEGDF 714
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2206 INIITKPGKPGPGLVEVSEHKGDKLKKPKPDGGEPEVSELYVERKFPDPTGIWLPVG-- 2263
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 715 KFEQNRNKRREDRTFLVGNLEIRERLRLNDVPESSRCRCVKRAYRSEFPLPSDIOGVV 774
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2264 -----RSDGE-----YVNDGLVGHDK--FR 2284
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 775 ISVINLEPRTGFLSNBRAMGRFDSVITGPGACVPAFCDDQSPDAYSAVYASLAGEELQ 834
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2285 VKAVNKE-----GESEPLET-----LGSIIAKDPF 2309
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 835 AVESSEKFNPNALIGVPOPLKLNLRDHDPRVKTAFOISMARPNSEESNGPI- 893
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2310 SVPTKRP-----GVPEP-----TDW-----TANKVELAMPKP--ASDGSPIQ 2344
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 894 -YAFENL-----RACEAPSAHPRFYQIEGDRDYNTVPPE 931
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2345 GIVEYKDKYSPLMERKALETNSPTPTATVQGLIEGNEYQFRVALNK 2391
| : : : : : | : : : : : | : : : : : | : : : : : |

```

```

RESULT 7
JH0771
protein-tyrosine kinase (EC 2.7.1.112), receptor type tyk precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 21-Jul-2000
C:Accession: JH0771
R:Horita, K.; Yagi, T.; Kohmura, N.; Tomooka, Y.; Ikawa, Y.; Aizawa, S.
Biochem. Biophys. Res. Commun. 189, 1747-1753, 1992
A:Title: A novel tyrosine kinase, tyk expressed in murine embryonic stem cells.
A:Reference number: JH0771; MUID:93129253
A:Accession: JH0771
A:Molecule type: mRNA
A:Residues: 1-1125 <HOR>
A:Cross-references: GB:D13738; NID:g220439; PIDN:BA02883.1; PID:g220440
C:Genetics:
A:Gene: tyk
C:Superfamily: protein-tyrosine kinase, receptor type tie; EGF homology; fibronectin
C:Keywords: Arp; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos
F:1-122/Domain: signal sequence #status predicted <SIG>
F:23-1125/Product: protein-tyrosine kinase, receptor type tyk #status predicted <MAT>
F:37-104/Domain: immunoglobulin homology <IM1>
F:135-137/Region: cell attachment (R-G-D) motif
F:212-252/Domain: EGF homology <EG1>
F:256-299/Domain: EGF homology <EG2>
F:303-341/Domain: EGF homology <EG3>
F:365-427/Domain: immunoglobulin homology <IM2>
F:448-526/Domain: fibronectin type III repeat homology <FN3A>
F:541-625/Domain: fibronectin type III repeat homology <FN3B>
F:638-720/Domain: fibronectin type III repeat homology <FN3C>
F:752-773/Domain: transmembrane #status predicted <TM>
F:832-1100/Domain: protein kinase homology <KIN>
F:831-839/Region: protein kinase Arp-binding motif
F:1140,138,400,439,465,550,556,649,691/Binding site: carbohydrate (Asn) (covalent) #st
F:856,873,965/Active site: Lys, Glu, Asp #status predicted

```

```

Query Match 2.4%; Score 148.5; DB 1; Length 1125;
Best Local Similarity 17.7%; Pred. No. 0.069;
Matches 174; Conservative 117; Mismatches 320; Indels 373; Gaps 40;

```

```

OY 10 SFLVEVTSVLGRQMTLQSVRRVOPGKNPISIFKPA--DTLESFG--EMTTLFNIDYFG 67
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 123 SFLPATLFTMTVDRGNVNISFKVLIKEDAVTYKNGSLHPLSARGMKYDLILEVHLRHA 182
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 68 KCDYERLDAIRYYGDRVCARLREARTD--WTPACSTQGVVNGSPRESFGLNREQ 124
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 183 QPDAGVYSARYIGNLFSAFTLIVRCEAQKGPDS----- 222
| : : : : : | : : : : : | : : : : : | : : : : : |
OY 125 RFGQCSNV-----TVRLCPGSLRNDTERIWSFPWSPWMSKCSAAC--GQTGVOTRTRIC 177
| : : : : : | : : : : : | : : : : : | : : : : : |

```

```

Db 223 RPTCTCKNNGVCHBDTGCICPFGFMGRTCEKACEPHTFGRCKERCSCGPEGCAASY-FC 281
178 LAENVSLCSEASEEGOHCMQDCTACDLTCMGVYMAC-----DACMCQ 222
282 LPDYG-CSCAT-----GWRGIQCNACPSGGYGPCKLRCHCTNEICDRFGGLCS 333
223 DFMHGAVALPFGAPASGAAYLLTKTPKLTQTDSDGRFRIPGLCPDGKSLKTKYKF 282
334 -----QMGGL-----QCEKGRPMPTQIIDLPHIEVNSKF 367
Db 283 APIYLMPKTSLKAAITKAETVRAETPYMMNPETKARRAGQSVLCCATGKPRP--D 339
368 NP-----CKASGMPLPPTSEE 383
340 KYFVYHNDLLDPSLYKHESKLVLRKIQHQAGYFCKAQSDAGAVSKVQLVTASDE 399
384 MTLVKGPGTYIQPNDFTYDF-----SALFTVNRV 415
400 TPCNPVPESTYLRPHDCFOATNSFYDVGRCVKTCAQGOONGIRCAVONCCGISK 459
416 LP-----PDS-----GVVVCSTNTVAGK----- 433
460 TEEHEIQCSGTLPTKAKECSCQCTETRSIVGRVSAADNGEPARFHHY-WGNS--- 515
434 -VEKFNISYKVLV-----EPLHAPNVIDTGHAFAI 463
516 -RVSMGTGKGFTHVPODTEFLVTFVDRLQKFNVTTKVLPENKGSAAVHEIKMLRK 574
464 INISSEPFYF--DGPFKSKLFFKPVNOAKTIEVT-----NEFTLNTYL 506
575 EPTLEAMEFNITLPLGEVVGEDPALEIDRSFVRONGEPIYGVK---ASVTFLLDP 630
507 EPRDYELCYQLARPG-----GGECHPFPVRRFTTASIGLPPR 546
631 NISATGAOTDLN-----FINDEGDTFPLRTYGMFSVDRDETSPLNAGKVVLD 684
547 GLSLPSSQALNLTWOPITNSD-----EFVEVERSLQTS---DQ 588
685 TOVKMPHISTVKLMSLN-----DTGLWEE-----GDF--KPEORRN 722
589 QNIKVPENLTVLLSNLVPREQYTVARAVNTKAQGESEELRAWLTDLIPROENIKIS 648
723 KREERTLVGNLEIRERLNLV-----PESRCFVYVRAVRSRFLPSDIOGVY 775
649 NITDSTAMVMTIVDGYSSISITIRYVQGNEDQHDIVKIKATVTOY---QLKG--- 701
776 SVINLEPRTGFLSNPRAMGRFDSVITGNGACVPAFCDD-----QSPDAYSAVYLA 830
702 -----LEBETTY-----HVDIFAENNIGSSNPAFSEHLRLTLPSPAS-----ADLGG 743
831 EELQAVSSKRFENNAIGV--QPYLKLKLYRTD-----HEDPRVAKTAFOI 876
744 GK-DATSHHWVWMDNFASPCLLAFILMLQKRAVQRRMAQAFQVNRPAVQFNGSL 802
877 SMAKPRNSAESNGPIYAFENLR 900
803 ALNRKAKNNDPITYIVLDMNDIK 826

```

RESULT 8
 T25061
 hypothetical protein T21B6.3 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T25061
 R:cottage, A.
 submitted to the EMBL Data Library, November 1995
 A:Reference number: Z19975
 A:Accession: T25061
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-788 <WIL>

A:Cross-references: EMBL:Z68011; PTDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3
 A:Experimental source: clone T21B6
 C:Genetics:
 A:Gene: CESP:T21B6.3
 A:Map position: X
 A:Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2

Query Match 2.3%; Score 147.5; DB 2; Length 788;
 Best Local Similarity 27.28; Pred. No. 0.048;
 Matches 55; Conservative 15; Mismatches 65; Indels 67; Gaps 11;

```

QY 46 PADLESPEGWTTLFNIDYPGKGDYERLDAIR-----YGGVRCARPLRL 93
Db 501 PCQWSEMCST-----CSASGSGQRER---TRCHLGTNRCEKDYESQCSAGPCPEW 554
QY 94 ARTDWTAGST-GQVYHGSFREGFWCL-----NREOR--PGNCSNYTVRFL 138
Db 555 SOMEWMGOCSTYCGQAVARQRT---CLGGVFQDHLCCGPKTEQRAACDGGPCS----- 604
QY 139 CPPESLRDRTERIMSPWSPMSKCSAAGOTGVOTRTRICLAENVSLCSEASEEGOHCMQ 198
Db 605 -----LMSFWQEMSTCSASC--SGMKRQRYC--QETDCQGFNESSQFCYGP 649
QY 199 DC-----TACDLTCPMGQ 211
Db 650 PCAEWTEMCWSSGSSKCGPGQ 671

```

RESULT 9
 T18856
 angio genesis inhibitor homolog - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T18856; T24653
 R:McMurray, A.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: Z19031
 A:Accession: T18856
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1444 <WIL>
 A:Cross-references: EMBL:Z50004; PTDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1
 A:Experimental source: clone C02B4
 R:McMurray, A.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: Z19917
 A:Accession: T24653
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1444 <WIL>
 A:Cross-references: EMBL:Z50006; PTDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1
 A:Experimental source: clone T07C5
 C:Genetics:
 A:Gene: CESP:C02B4.1
 A:Map position: X
 A:Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3;

Query Match 2.3%; Score 147; DB 2; Length 1444;
 Best Local Similarity 28.98; Pred. No. 0.13; 64; Indels 22; Gaps 8;
 Matches 44; Conservative 22; Mismatches 64; Indels 22; Gaps 8;

```

QY 86 CARPLRLAART-----TDWTP-AGSTGVYHGSFREGFWCLNREORPGQNCSTYVR-F 137
Db 1297 CFPDPARHLRDGMSWTMSDWTPCASGSGFQVQRNDS---CSSPFRKGQSGSGLAHQNS 1353
QY 138 LCPPESLRDRTERIMSPWSPMSKCSAAGOTGVOTRTRICLAENVSLCSEASEEGOHCMG 197
Db 1354 LCPDLPACDHESDGEWSAMNEMSGCMGNC--IGTRTVRAVSPV-----SDGQGPFCG 1406
QY 198 QDCTACDLT-CPMGVYNAODDCAQCFMFMH 228

```


Db 1407 R---SSEITECRQSPSTALCSSFITSHLADG 1435

RESULT 10

A32579 neuroglian - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A32579

R:Blaber, A.J.; Snow, P.M.; Hortsch, M.; Patel, N.H.; Jacobs, J.R.; Traquina, Z.R.; Schi Cell 59, 447-460, 1989

A:Title: Drosophila neuroglian: a member of the immunoglobulin superfamily with extensiv

A:Reference number: A32579; MUID:90030418

A:Accession: A32579

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1239 <BIE>

A:Cross-references: GB:M8231; NID:g157998; PIDN:AAA28728.1; PID:g157999

A:Note: the authors translated the codon TAT for residue 1234 as Thr and AAA for residue

C:Genetics:

A:Gene: FlyBase:Flyg

A:Cross-references: FlyBase:FBgn0002968

C:Superfamily: neutral cell adhesion molecule LI; fibronectin type III repeat homology; I

C:Keywords: alternative splicing; cell adhesion; duplication; membrane protein

F:353-412/Domain: immunoglobulin homology <IMM1>

F:446-502/Domain: immunoglobulin homology <IMM2>

F:535-596/Domain: immunoglobulin homology <IMM3>

Query Match 2.38; Score 146; DB 1; Length 1239;

Best Local Similarity 20.38; Pred. No. 0.12;

Matches 149; Conservative 99; Mismatches 277; Indels 208; Gaps 41;

304 VRAEPTVAMNPEKARAGOSVSLCKATGKPRPKYFWYHNDLDPSTLYKHSK--L 361

427 VQAEPTTSEAPAAVSTVDGKRVITKCRVNGSPK-LVKWILASWMLTGRNVAQNDL 485

362 VLKRLQHQAGEYPCKAQSDAGKVAQLIVTASDETPCPVPESTYLIRPHCFQNA 421

486 EIQDVTSDAGYTCYACNKRGEIQAD-GSLV--KEHTRITQEQNTYEA---AGQSA 538

422 T---NSFYDVGRCPCVTKCA-GQ-----ODNGIRCDAVONCGISTEEREICSGYT 471

539 TPCREHNDLLEIFIDMKKQGSIDFPAQPRFVKTND---NSLTIAKT--MELDSGEYT 593

472 LPTKVAECSCORCTETSTIVR-----GRVSAADNGEPMRGHYMGNSRVSMGTG 522

594 C---VARRLEDAFARMLIVQDVNPAPKLTGTCQADKAE---IHWEGQGNRSPIILHY 647

523 KGTf-TLHVP-----QDTERLYTFYDRLOKFPV-N-TTKVLPRKKGSA--VFHEIKML 571

648 TIOENTSTTPASMDAAVYKVPNTDSSFYVQMSPMANLYFRVIAFNKIGASPSASDSCT 707

572 RR-----KEPTLEAMENIIPLEGVEGDPMAELE-----IPSRSF-- 608

708 TOPDVPFKNPNVVOGGE--PNNLYISWTMPETLEHAPNPHYVSKKRIIPAAWENN 765

609 ----YRQNG-----EPYIGKVAASYTELDPNISTATAQTDLNFINDGDTFPLRTYGM 659

766 NIFDRQNNIYIADQPTFVKYLIKVAINDRGESVVAEEV---VGSGEDRPIDAPLN 821

660 FSVDRDVTSE-----PLNKGKVAVHLDSTQYKMPHEHISTYKWLMSLNDTGLMEE 711

822 FTM---RQITSTSGYMAWTPVSESVKGRHKYKX-----IQTWT-----ENE 861

712 GDFKEENOR--RNKREDTFLVGNLEIRERLLNDVPESSRCFVAVRSEERFLPSBO 769

862 GE---EGAREIHVKGDTHNALVTOFK-----PDSKN-VARILAYNR----- 899

770 IQGVVIVINLEPRTGLSNPRAMGRFDSVITGPNACVPAFCDDQSPDAY---SAYVL 825

900 -----FNGPSPASVVIDPDT---PEGVPSPV---QGLDAYVPLGSSAFWL 935

Oy 826 -----ASLAGELOAVESSPKFNPAIGVPOPLYLNKLYRRDDEDPKVKTAFOI 876

Db 936 HMKKRLYPNGKTLTGKRIYEE-----VKSEYGERREYDPHTDPVTR-----M 980

Oy 877 SMAKRPNS-----AESNGPIYAFE--NLRACEAPPSAAHFRYQTEGD----R 921

Db 981 KWAGLKPNKSKYRISTATATTKMGEGSEHYIEKTTLDKAVAVAPATPSFWEQLPNDGLAK 1040

Oy 922 YDYNVTPNEDDP 934

Db 1041 FRINWLPSTEGHP 1053

RESULT 11

S68235 myosin-light-chain kinase (EC 2.7.1.117), 210K, nonmuscle - chicken

N:Contains: myosin-light-chain kinase, 108K, smooth muscle; telokin

C:Species: Gallus gallus (chicken)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S68235; A37099; B44389; S28227; S78216; A35093; A25810; S11652

R:Materson, D.M.; Collinge, M.; Lukas, T.J.; Van Eldik, L.J.; Birukov, K.G.; Stepano

FEBS Lett. 373, 217-220, 1993

A:Title: Multiple gene products are produced from a novel protein kinase transcriptio

A:Reference number: S68235; MUID:96033976

A:Accession: S68235

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1906 <WAT>

A:Cross-references: EMBL:X52876; NID:g992992; PIDN:CAA37056.1; PID:g992993

R:Shomaker, M.O.; Lau, W.; Shattuck, R.L.; Kwiatkowski, A.P.; Matrisian, P.E.; Guert

J. Cell Biol. 111, 1107-1125, 1990

A:Title: Use of DNA sequence and mutant analyses and antisense oligodeoxynucleotides

activity.

A:Reference number: A37099; MUID:90361738

A:Accession: A37099

A:Molecule type: mRNA

A:Residues: 649-1906 <SHO>

A:Cross-references: EMBL:X52876

R:Collinge, M.; Matrisian, P.E.; Zimmer, W.E.; Shattuck, R.L.; Lukas, T.J.; Van Eldik

Mol. Cell. Biol. 12, 2359-2371, 1992

A:Title: Structure and expression of a calcium-binding protein gene contained within

A:Reference number: A44389; MUID:92236611

A:Accession: B44389

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1750-1906 <COL>

A:Cross-references: GB:M8284; NID:g212237; PIDN:AAB53767.1; PID:g212238

A:Accession: A44389

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1750-1906 <CO2>

A:Cross-references: GB:M8283; NID:g211371; PIDN:AAA48647.1; PID:g211372

R:Yoshikai, S.I.; Ikebe, M.

Arch. Biochem. Biophys. 299, 242-247, 1992

A:Title: Molecular cloning of the chicken gizzard telokin gene and cDNA.

A:Reference number: S28227; MUID:93073972

A:Accession: S28227

A:Molecule type: mRNA

A:Residues: 1750-1906 <YOS>

A:Cross-references: EMBL:M96655; NID:g212744; PIDN:AAA49083.1; PID:g212745

A:Accession: S78216

A:Molecule type: DNA

A:Residues: 1750-1906 <YOW>

A:Cross-references: EMBL:M96987

R:Olson, N.J.; Pearson, R.B.; Needleman, D.S.; Hurvitz, M.Y.; Kemp, B.E.; Means, A.R.

Proc. Natl. Acad. Sci. U.S.A. 87, 2284-2288, 1990

A:Title: Regulatory and structural motifs of chicken gizzard myosin light chain kinas

A:Reference number: A35093; MUID:90192792

A:Accession: A35093

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 935-1438, 'Q', 1440-1906 <OLS>

A:Cross-references: GB:M31048; NID:g212660; PIDN:AAA49069.1; PID:g212661

R;Guerrero Jr., V.; Russo, M.A.; Olson, N.J.; Putkey, J.A.; Means, A.R.
 Biochemistry 25, 8372-8381, 1986
 A:Title: Domain organization of chicken gizzard myosin light chain kinase deduced from a
 A:Reference number: A25810; MUID:87157587
 A:Accession: A25810
 A:Molecule type: mRNA
 A:Residues: 1258-1438, 'O', 1440-1906 <GUE>
 C:Genetics:
 A:Introns: 1735/3; 1779/1; 1819/1
 C:Superfamily: myosin-light-chain kinase, nonmuscle; fibronectin type III repeat homolog
 C:Keywords: alternative initiators; ATP; calmodulin binding; phosphoprotein; phosphotran
 F:542-599/Domain: immunoglobulin homology <IMM1>
 F:935-1906/Product: myosin-light-chain kinase, 108k, smooth muscle (from 5.5kb transcript
 F:1098-1158/Domain: immunoglobulin homology <IMM2>
 F:1451-1708/Domain: protein kinase homology <KIN>
 F:1459-1467/Region: protein kinase ATP-binding motif
 F:1750-1906/Product: telokin (kinase-related protein KRP) (from 2.7 kb transcript)
 F:1808-1869/Domain: immunoglobulin homology <IMM3>

Query Match

2.3%; Score 145; DB 1; Length 1906;
 Best Local Similarity 19.3%; Pred. No. 0.26;

Matches 179; Conservative 109; Mismatches 305; Indels 334; Gaps 45;

QY 140 PPSLRDTERIMSPWSPKSCASACQOTGYOTRTRICLAEMVSLGSEASE-----GQ 193
 DB 1111 PPAVS-----SWTLDK-----AIKSSKSTIVISOGELICSLTEKWPEDGGE 1153
 QY 194 H-CMGODCTACDLTCPMGOVNADCDACMCDFMLHGAVALPGAPASGAAIYLLTKTPKL 252
 DB 1154 YKCAENNAAG-----KAEC-AC---KVLVEDTSSIRAKPAKAKTKKPPVTTLPV 1199
 QY 253 LQOTDSGRFRIPGLCPDGKSLIKTRKVPAPVLTMPKTSLKATIKAFVRAETPYMV 312
 DB 1200 LSTESSE-----ATVKKKP-----APKTPPKKAT-----PQIT 1228
 QY 313 MNPETKARRAGOSYSLCCKATGKRPDKYFMYHNDTLDDSLY-----KHESLYLRKL 366
 DB 1229 QPFDKRVKRGESVELPAKAVG--TAPITCTWMKFRKQIQEVEYKIKENAESKLTTSST 1287
 QY 367 QOHQAGEYFCKAOSDAGAVSKVAQLIVTASDE---TPC-NPVESYLIRLPHDFONAT 422
 DB 1288 KOEHCQGYTLIVENKLGSRQAQVNLTVVDKPPDPAITPCASDIRSSL-----T 1336
 QY 423 NSFY---YDVGR-----PVKTCAG-----QDNGICRDVAQNC 454
 DB 1337 LSWGSSYDGSANVOSTYVEIWNVDNKTDLTTCRSTSFNVOLDQDREKFRVRAVY 1396
 QY 455 CGIS-----KTEEREIOCGSYTLPTKVAKCSQORCTETSIYRGVRSADNG 502
 DB 1397 YGISEPQSEEVYKVGKQEBELKEEPAELSDDEGKETEVNRYVTYN--TEOKVSDVYNI 1455
 QY 503 EPM---RFQHVYVGNRSVMTGKGT-----TLHVQDTERLV 538
 DB 1456 EHRIGSKFQGOVRLVBEKTKGKVAAGKFKAYSAKKEKNIRDEISIMCLHHR-----L 1510
 QY 539 LTFVDRLOKRVNTTKVLPFNKKSGSAVFEIKMLRKKEPITLEAMETNIRPLGVEVGEDPM 598
 DB 1511 VQCVDAPEEKANIVMLEM--VSGGELFERI-----IDED-- 1543
 QY 599 AELEIPR---SFYRQ---NEPYIGKASVTLDPNINISTATAQDILFINDEGDTFP 653
 DB 1544 -FELERECIKYMKROISEVEYIHKQGIVHLDPKINIMVNTKTSIKLID----- 1594
 QY 654 LRTVGMVSVPFRVETSEPLNAGKVVKHLDSTQVKMEHS-----TVKLMS----- 700
 DB 1595 -----FGLARRLE-----SAGSLKVLFGTPEFVAVPEVINYEPIGETDMWSIGVICYI 1642
 QY 701 -----LNDPTGLWEE-----GDFKFNQRRNK--REDRTFLVGNLEIRBRRLFNLDVPE 748
 DB 1643 LVSGLSLSPFGMDNDNETLANVTSATWDFDPAFDEISDADDFISN-----LLKKMKMS 1695
 QY 749 SRRC-----FVKVRAVRSERLPSQIQGVIVSVINLEPRTGFLSNPR 791

DB 1696 RLNTQCLQHPWLQDKTRNMEAKKLSKDRMKKYARRKWO-----KFGHA--VR 1742
 QY 792 AMGFEDS--YITSPNGACVAFACDDQSPDAYSAVYLAISLAGELOAVESSPKRPNMIGV 849
 DB 1743 AIGRLSSMAWISGMSGR-----KASGSSP----- 1766
 QY 850 POPYLNKLYNRRTDDEHPRYKTAFOJISMAKPRENSAESNGPIYAFENIRACEAPPSA 909
 DB 1767 -----TSPINDKVENDEAPLEVAEKEPRHKP-----YFKTITLDEVEYEGSA 1810
 QY 910 AHFRFYIEGDY--DYNTVPENEDDPM 935
 DB 1811 ARFD-CRIEG--YPDPEVMYKDDQPV 1834

RESULT 12

JN0583
 myosin-light-chain kinase (EC 2.7.1.117) 155k protein - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 20-Apr-2000

C:Accession: JN0583; P0490

R;Kobayashi, H.; Inoue, A.; Mikawa, T.; Kuwayama, H.; Hotta, Y.; Masaki, T.; Ebashi,

J. Biochem. 112, 786-791, 1992

A:Title: Isolation of cDNA for bovine stomach 155kDa protein exhibiting myosin light

A:Reference number: JN0583; MUID:93203148

A:Accession: JN0583

A:Molecule type: mRNA

A:Residues: 1-1176 <KOB>

A:Cross-references: GB:557131; NID:g298638; PIDN:AAB25794.1; PID:g298639

A:Experimental source: stomach

A:Accession: P0490

A:Molecule type: protein

A:Residues: 44-55;721-728;828-851;1002-1019 <KO2>

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; fibronectin type I

C:Keywords: ATP; calmodulin binding; phosphotransferase

F:370-430/Domain: immunoglobulin homology <IMM1>

F:723-980/Domain: protein kinase homology <KIN>

F:731-739/Region: protein kinase ATP-binding motif

F:1083-1144/Domain: immunoglobulin homology <IMM2>

Query Match

2.3%; Score 144.5; DB 2; Length 1176;
 Best Local Similarity 20.3%; Pred. No. 0.14;

Matches 179; Conservative 115; Mismatches 306; Indels 281; Gaps 47;

QY 166 GQGVQVOTRTRICLAEMVSLGSEASEGQHCMDGOD-----CTACDLTCPMGOVNADCDACM 220
 DB 393 GKT-LKTRKTVLSQESLSYSTIEK---ALPEDRGILKYVAKN---SAGQAESGCV-- 443
 QY 221 QDFMLHGAVALPGAPASGAATYLLTKTPKLLTQTDSDGRFRIPGLCPDGKSLIKITV 280
 DB 444 -----TVDP--DAPTSNA-----KADEMKAARRKSSLPYLGTESD-----ATV 482
 QY 281 KPAIVLTMRKTSLSKAATTAEFVRAETPYVMANPETKARRAGOSYSLCCATGKPRPK 340
 DB 483 KKPV---AKTTPKAM-----PQIIQFPEDQKRVAGESVELFKVAG--TQPIR 528
 QY 341 YFWV--HNDTLDDSLYKHE-----SKLYLRKLOHQAGEFCKAOSDAGAVSKVAQLIV 394
 DB 529 CTWKKFRKQIQDSEBHKIVENSEGSKLTITRAAOEHGCGCTLLVENKLGSRQAQVNLTV 588
 QY 395 TASDE---TPC-NPVESYLIRLPHDFONATNSFY---YDVGRCPVKTGAGQDNGIRC 447
 DB 589 DKDPDPACTGASDIRSSL-----TISWIGSSIDGG-----SAVQISVET 630
 QY 448 RDAVQ-----NCCGIS-----KTEEREIOCGSYTLPTKVAKCSQORCTET 488
 DB 631 WDSVDKTKWELATCRSTSFNVODLLPDREYKFRVRAINVCSTSEPSQESRLTALGEKPEE 690
 QY 489 RSTVIRGVSAADNGEPARFQHVYVGNRSVMTGKGTGFTLHVQDTERLVLTVD---RL 545
 DB 691 EPDEVEVSDDEREP-----EVDYR--TVTV-----NTEOKVSDFYDIEERL 731

F:151,316,330,457,584,710,1069/Binding site: carbohydrate (Asn) (covalent) #status predi
 F:167-226/Disulfide bonds: #status predicted
 F:266,270/Disulfide bonds: interchain #status predicted
 F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 2.3%; Score 143; DB 1; Length 1172;
 Best Local Similarity 20.8%; Pred. No. 0.18;
 Matches 105; Conservative 56; Mismatches 189; Indels 154; Gaps 29;

QY 86 CAPRLRLAET-----TDMTASGTVGVHSPRGFMCLNREQRPGQ 128
 DB 360 CASPSEYEGCCSCSLHSDVEBWSMSPMAEMTQCSVT-----CGSGTQGRGR 406
 QY 129 NCSNYTRFLCP-----PGSLRDRTERI-----WSPWSPWKSACAGQTVQTRICL 178
 DB 407 SCQVTSNTCLGPSIQTRACSLSKDFTRIQDDGSHMSWSPSSCVTCG-VGNITRIQLCN 465
 QY 179 AEWVSL---CSEASEGQHCMGQDC-----TACDLTCPMG--QVNAACD--- 217
 DB 466 SPVPOMGKMKSGRETAKOGAPCPIDGRWSPWSPWSACTVTCAGGIRETRVCNSPE 525
 QY 218 -----AC-----MC--ODFMLHGVSLP--GAPASGAIVLTKTPKLLTQTDSD 259
 DB 526 PQYGGKACVGDVQEROMCKNRKSCPVDCLSNCPGPGACS-----SFPD 569
 QY 260 GRRFRIPLCP-----DGKSLIKITKVFAPIVLTPMTSLKAATIKAEFVRAETP--YVW 312
 DB 570 GSWSC--GFCVGFNGNTHCEDDELCALVPDICEFSTKVPKVCNTQPGFHLPCPPRYRG 628
 QY 313 MNPETARRRAGQSVSLCCKATGKPRDKYFYINNTLLDPSLYKHESKLVLRKLQOHQAG 372
 DB 629 NOPVGVGLEAKTEKQVCEPE-NPCKDK---TNN-----CHKHAECIYLGHFSDPM-- 675
 QY 373 EYFCKAQSADGAVKSKAQLIVTASDETPCNPVPESYLIRLPHDCFQNAFNSFYDYV-GR 431
 DB 676 -YKCECQTG-----YAGDELICG--EDSDDGMPLNLYCATNATYHCLIKDN 719
 QY 432 CPVKTCAGQDNGIRCDAYVONCC-----GISKTEEREIQCSTYLLPTVAKESG--- 481
 DB 720 CPHLPMNSGGDFD---KDGIGDADDDDDNDGY--TDEKD-NCQLLENPRQADYDKDEVG 773
 QY 482 --CORCTRSYIRGRVSAADNGE 503
 DB 774 .DRCDNCPYVHN--PAQIDTDNNGE 795

RESULT 15

T339322
 Hypothetical protein Y8A9A.2 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T339322
 R:Courtney, L.; Langston, Y.; Drone, K.; Mead, K.
 submitted to the EMBL Data Library, February 1999
 A:Description: The sequence of *C. elegans* cosmid Y8A9A.
 A:Reference number: Z21439
 A:Accession: T339322
 A:Status: preliminary;
 A:Molecule type: DNA
 A:Residues: 1-1360 <CON>
 A:Cross-references: EMBL:AF125461; PIDN:AD12852.1; GSPDB:GN00020; CESP:Y8A9A.2
 A:Experimental source: strain Bristol N2; clone Y8A9A
 C:Genetics:
 A:Gene: CESP:Y8A9A.2
 A:Map position: 2
 A:introns: 43/3; 69/3; 134/3; 174/1; 604/2; 643/3; 718/3; 787/3; 848/2; 944/2; 1044/2; 1

Query Match 2.3%; Score 143; DB 2; Length 1360;
 Best Local Similarity 26.1%; Pred. No. 0.22;
 Matches 61; Conservative 16; Mismatches 95; Indels 62; Gaps 11;

QY 54 GEMTTLFNI-DYRGGKG-DYERLDAIRFYVG-----DRVCARPLRLAETDMT--- 100
 DB 1108 GEMVSVSGCNDTCGSCGQETRRKRKCLSLQYGCACGTGNATDTVCASSVCLFPRTSCTGF 1167
 QY 101 --PAGSTGVVHSPRGFMCLNREQRPGQNCNITYRFLCPPGSLRDRTERIWSWSPW 158
 DB 1168 KKNVNTIGRTFYCGPLPVPVAFNPEQ---TTC-----CDP-----EXTGLMNMGAW 1211
 QY 159 SKCSAAGQTVQTRFTRICLAEWVSL-CSEASEGQHCMGQDCCTACDLTCPMGVNACD 217
 DB 1212 TTCGATCGGCGTQTRSKTCASAPYGCPCCTGDLTETOSCAKQVCTTGACCCAGKFAVATGYD 1271
 QY 218 -ACMCD-----FMLHGV-----SLPGAPASGA 241
 DB 1272 GAQYCDNTPREVCTGTWTEWATLEGAVCNDPCGNCGLIPTSRYCPSGCGCSGA 1325

Search completed: October 9, 2001, 11:50:42
 Job time: 227 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2001, 11:50:15 ; Search time 14.83 Seconds

(without alignments)
2734.895 Million cell updates/sec

Title: US-09-609-383-2

Perfect score: 6317
Sequence: 1 MVTGKAWFSEFLVLEVTSL.....QSGVVASLRFPVRAQOPLIN 1184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	170	2.7	890	1	AT88_HUMAN
2	169	2.7	1170	1	TSPL_MOUSE
3	166	2.6	1170	1	TSPL_BOVIN
4	162	2.6	1170	1	TSPL_BOVIN
5	161	2.5	1694	1	SN_MOUSE
6	160	2.5	905	1	AT88_MOUSE
7	159.5	2.5	2569	1	LMA3_MOUSE
8	157	2.5	1170	1	TSPL_HUMAN
9	151.5	2.4	4393	1	PGBM_HUMAN
10	146	2.3	1239	1	NRG_DROME
11	145	2.3	1173	1	TSPL_XENLA
12	145	2.3	1906	1	KMLS_CHICK
13	144.5	2.3	1176	1	KMLS_BOVIN
14	144.5	2.3	1914	1	KMLS_HUMAN
15	143	2.3	1172	1	TSPL_HUMAN
16	142.5	2.3	4544	1	LPL_HUMAN
17	142	2.2	1074	1	SM5A_HUMAN
18	140.5	2.2	2481	1	UN52_CAEEL
19	139.5	2.2	1056	1	MUC5_HUMAN
20	139	2.2	1122	1	TIE2_MOUSE
21	139	2.2	1522	1	BAL1_HUMAN
22	137.5	2.2	1493	1	NR01_MOUSE
23	136	2.2	2907	1	FBN2_MOUSE
24	135	2.1	1147	1	KMLS_RABIT
25	135	2.1	2911	1	FBN2_HUMAN
26	134.5	2.1	1450	1	MPSE_CHICK
27	133.5	2.1	610	1	TOH2_CAEEL
28	132	2.1	1178	1	TSPL_CHICK
29	132	2.1	1211	1	AT52_HUMAN
30	131.5	2.1	3707	1	PGBM_MOUSE
31	130.5	2.1	1125	1	TIE2_BOVIN
32	130	2.1	470	1	PROP_CAVPO
33	130	2.1	867	1	SSPO_BOVIN

ALIGNMENTS

RESULT	ID	AT88_HUMAN	STANDARD	PRT	890 AA.
AC	Q9UP79	Q9NZS0			
DT	01-OCR-2000	(Rel. 40, Created)			
DT	01-OCR-2000	(Rel. 40, Last sequence update)			
DT	01-OCR-2000	(Rel. 40, Last annotation update)			
DE	ADAM-TS 8 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE WITH THROMBOSPONDIN MOTIFS 8) (ADAMTS-8) (ADAM-TS8) (METH-2) (METH-8).				
GN	ADAMTS8 OR METH2				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Lung;				
RX	MEDLINE=99367466; PubMed=10438512;				
RA	Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Olkenus S., Lombardo M., Iruela-Arispe M.L.;				
RT	"METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity.";				
RL	J. Biol. Chem. 274:23349-23357(1999).				
RN	[2]				
RP	SEQUENCE OF 195-440 FROM N.A.				
RC	MEDLINE=20079168; PubMed=10610729;				
RA	Georgiadis K.E., Hirohata S., Seidlin M.F., Apre S.S.;				
RT	"ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on mouse chromosome 9 and human chromosome 11.";				
RL	Genomics 62:312-315(1999).				
CC	-1- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.				
CC	-1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).				
CC	-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND KIDNEY.				
CC	-1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.				
CC	-1- PFM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC METALLOPROTEINASE); ALSO KNOWN AS THE REPEROLYSIN SUBFAMILY.				
CC	-1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.				
CC	-1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AF060153; AAD48081.1; -				
DR	EMBL; AF175283; AAF25806.1; -				

DR HSSP: P34179; 11AG.
 DR MIM: 605175; -.
 DR InterPro: IPR000130; -.
 DR InterPro: IPR000884; -.
 DR InterPro: IPR001590; -.
 DR InterPro: IPR002870; -.
 DR Pfam: PF000090; tsp.1; 2.
 DR Pfam: PF01421; Repolysin; 1.
 DR Pfam: PF01562; Pep_M12B_propep; 1.
 DR PROSITE: PS50215; ADAM_MEPRO; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 DR PROSITE: PS50092; TSP1; 2.
 DR PROSITE: PS00427; DISINTEGRINS; FALSE_NEG.
 DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 DR Repeat; Extracellular matrix; Heparin-binding.
 FT SIGNAL 1 27
 FT PROPEP 28 214
 FT CHAIN 215 890
 FT METAL 364 364
 FT ACT_SITE 365 365
 FT METAL 368 368
 FT METAL 374 374
 FT DOMAIN 439 526
 FT DOMAIN 527 583
 FT DOMAIN 584 690
 FT DOMAIN 691 832
 FT DOMAIN 833 890
 FT DOMAIN 202 205
 FT CARBOHYD 345 345
 FT CARBOHYD 401 401
 FT CARBOHYD 466 466
 FT CARBOHYD 491 491
 FT CARBOHYD 600 600
 FT CONFLICT 195 195
 FT CONFLICT 413 440
 SEQUENCE 890 AA; 96671 MW; 57070EB03D5739D3 CRC64;
 IHEKYLCKGCHGDCILDAPGALPLPTGL -> FSGCHLQGM

Query Match 2.7%; Score 170; DB 1; Length 890;
 Best Local Similarity 21.4%; Pred. No. 0.00044;
 Matches 110; Conservative 49; Mismatches 224; Indels 132; Gaps 25;

QY 50 LESPEWTLFNDYDGGKDYERLDAIRRYGD--RVCAKPLRLKARTDWTDPAGSTGQ 107
 DB 427 LDAPGALPL-PTGLPGRMALYOLDQOCROIFPDRFHCNNTSAOVCAQLOMHTDGAEP 485
 QY 108 VVHSGPREGFWCLNREORPGONCSNTYVRLCPGSLRDRTERI-----WSPWSPWSK 160
 DB 486 LGHTKXGSLPMWAGTFCGPHLCSBGS-----CLP---EEVEVRPKRYVDGGMAPWGPWGE 538
 QY 161 CSAAGQOTGVQTRTRICLAEMVSLCSEASEGQCMG-----ODCTACDLTCPMGOVNAD 215
 DB 539 CSRTCG-GGVQFHSRCKD-----PEPQNGRHYCLGRRAKYQSCHTEE--CPPD--GRS 587
 QY 216 CDKCMODEPLHNAVSLPGCAPASGAAYILITPKLILTOTDSDGRIRPGLC-PDGKSI 274
 DB 588 FREQQCEKNAYVYTDMDG-----LLQWPKYAGVSPRD---RCKLIRFARRSRSE 635
 QY 275 LKTKKFAPIVLTMPKTSIKATIRAEFRATPYVMNPE--FKARRAGQSVSLCKCA 332
 DB 636 FKVEFAKVIDGLCPET--LAICVAGQCVKACGDHVDSPRLDKGCVGGGNGNCRKY 693
 QY 333 TGNPRDKFWIIND-----TLID-----PSLYKHESKVLKRLQOHQAGEYFCKA 378
 DB 694 SGLTFTNY--GYNDIVTTPAGATNIDVQRSHPGVQNDGNALYATAD---GQYLING 747
 QY 379 QSDAGAVK-----KYAQLIVASDETPONPPESTLIL---PHCF-----418
 DB 748 NLAISAEODILVKTILKSGSIAFLERLQSRPLPEPLTVGLTVPGGVPPKVKYTF 807
 QY 419 -----ONATNSFY-----DVGRCFVKTCAGQODNGIRRD---A 450

DB 808 FVNDVDFSMOSSKERATTNIIOPLLHAQWVLDWSECSTCGAGMORTVECRDPSGQA 867
 QY 451 VONCCGISKTEEREIOCGGYTLPTFKVAKESCQRC 485
 DB 868 SATCNKALKRPD-----AKPCESQLC 888
 RESULT 2
 TSP1_MOUSE
 ID TSP1_MOUSE STANDARD; PRT; 1170 AA.
 AC P35441;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 1 PRECURSOR.
 GN THBS1 OR TSP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92128941; PubMed=1774063;
 RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,
 RA Jenkins N.A.;
 RL "Characterization of the murine thrombospondin gene.";
 RL Genomics 11:587-600(1991).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92147683; PubMed=1371115;
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 RA Dixit V.M.;
 RL "Characterization of mouse thrombospondin 2 sequence and expression
 RL during cell growth and development.";
 RL J. Biol. Chem. 267:3274-3281(1992).
 [3]
 RP SEQUENCE OF 1-490 FROM N.A.
 RX MEDLINE=9037546; PubMed=2398070;
 RA Bornstein P., Alf D., Devarayalu S., Framson P., Li P.;
 RL "Characterization of the mouse thrombospondin gene and evaluation of
 RL the role of the first intron in human gene expression.";
 RL J. Biol. Chem. 265:16691-16698(1990).
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 CC LAMININ AND TYPE V COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions, as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M62470; AAA50611.1; -
 DR EMBL: M62450; AAA50611.1; JOINED.
 DR EMBL: M62451; AAA50611.1; JOINED.
 DR EMBL: M62452; AAA50611.1; JOINED.
 DR EMBL: M62453; AAA50611.1; JOINED.
 DR EMBL: M62454; AAA50611.1; JOINED.
 DR EMBL: M62455; AAA50611.1; JOINED.
 DR EMBL: M62456; AAA50611.1; JOINED.
 DR EMBL: M62457; AAA50611.1; JOINED.
 DR EMBL: M62458; AAA50611.1; JOINED.
 DR EMBL: M62459; AAA50611.1; JOINED.
 DR EMBL: M62460; AAA50611.1; JOINED.
 DR EMBL: M62461; AAA50611.1; JOINED.

DR	EMBL	M62463	AAA50611.1	JOINED.	
DR	EMBL	M62463	AAA50611.1	JOINED.	
DR	EMBL	M62464	AAA50611.1	JOINED.	
DR	EMBL	M62465	AAA50611.1	JOINED.	
DR	EMBL	M62466	AAA50611.1	JOINED.	
DR	EMBL	M62467	AAA50611.1	JOINED.	
DR	EMBL	M62468	AAA50611.1	JOINED.	
DR	EMBL	M62469	AAA50611.1	JOINED.	
DR	EMBL	M81276	AAA53063.1	-	
DR	EMBL	J05606	AAA40431.1	-	
DR	EMBL	J05605	AAA40431.1	JOINED.	
DR	PIR	A40558	A40558		
DR	PIR	B42587	B42587		
DR	PIR	A37905	A37905		
DR	HSSP	P35355	LEMO.		
DR	MGD	MGI:98737	Thbs1.		
DR	InterPro	IPR000561	-		
DR	InterPro	IPR000884	-		
DR	InterPro	IPR001007	-		
DR	Pfam	PF00008	EGF_2	3.	
DR	Pfam	PF00090	tsp_1	3.	
DR	Pfam	PF00093	wvc	1.	
DR	PROSITE	PS00022	EGF_1	FALSE_NEG.	
DR	PROSITE	PS0186	EGF_2	1.	
DR	PROSITE	PSS0092	TSP1	3.	
DR	PROSITE	PS01208	WVC	1.	
RW	glycoprotein	Cell adhesion	Calcium-binding; Heparin-binding; Repeat,		
KW	EGF-like domain	Signal.			
FT	SIGNAL	1	18	POTENTIAL.	
FT	CHAIN	19	1170	THROMBOSPONDIN 1.	
FT	DOMAIN	19	232	HEPARIN-BINDING (POTENTIAL).	
FT	DOMAIN	316	373	WVC.	
FT	DOMAIN	379	430	TSP TYPE-1 1.	
FT	DOMAIN	435	491	TSP TYPE-1 2.	
FT	DOMAIN	492	548	TSP TYPE-3 1.	
FT	DOMAIN	549	587	EGF-LIKE 1.	
FT	DOMAIN	588	645	EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).	
FT	DOMAIN	646	690	EGF-LIKE 3.	
FT	DOMAIN	723	758	TSP TYPE-3 1.	
FT	DOMAIN	759	781	TSP TYPE-3 2.	
FT	DOMAIN	782	817	TSP TYPE-3 3.	
FT	DOMAIN	818	840	TSP TYPE-3 4.	
FT	DOMAIN	841	878	TSP TYPE-3 5.	
FT	DOMAIN	879	914	TSP TYPE-3 6.	
FT	DOMAIN	915	950	TSP TYPE-3 7.	
FT	DOMAIN	951	1170	C-TERMINAL.	
FT	SITE	926	928	CELL ATTACHMENT SITE (POTENTIAL).	
FT	DISULFID	270	270	INTERCHAIN (PROBABLE).	
FT	DISULFID	274	274	INTERCHAIN (PROBABLE).	
FT	DISULFID	551	562	BY SIMILARITY.	
FT	DISULFID	556	572	BY SIMILARITY.	
FT	DISULFID	575	586	BY SIMILARITY.	
FT	DISULFID	592	618	BY SIMILARITY.	
FT	DISULFID	599	607	BY SIMILARITY.	
FT	DISULFID	620	644	BY SIMILARITY.	
FT	DISULFID	650	663	BY SIMILARITY.	
FT	DISULFID	657	676	BY SIMILARITY.	
FT	DISULFID	678	689	BY SIMILARITY.	
FT	CARBOHYD	248	248	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	360	360	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	708	708	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	1067	1067	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CONFLICT	1025	1025	F -> L (IN AAA53063).	
SO	SEQUENCE	1170	AA: 129646	MM: 04435493615E7F06	CRC64:

```

Query Match      2.7%; Score 169; DB 1; Length 1170;
Best Local Similarity 20.6%; Pred No. 0.00076;
Matches 113; Conservative 38; Mismatches 151; Indels 246; Gaps 28;

QY 10 SFLVLEVTSVLCGRQTMLTQSVRRVQPGK-----KNPSIEAKPADLTSPGHWIT---- 58
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::| 111
Db 278 SSMVLEKLRIRIVITLQDSIRKRVKEENNELVSELRPRPCCHNGVQYNNNEKMTWDSCT 337

```

[illegible]

```

RESULT      3
TSPL_BOVIN  STANDARD;          PRI;    1170 AA.
ID          TSPL_BOVIN
AC          Q28178; Q28179;
DT          01-NOV-1997 (Rel. 35, Created)
DT          01-OCT-2000 (Rel. 40, Last sequence update)
DT          01-OCT-2000 (Rel. 40, Last annotation update)
DE          THROMBOSPONDIN 1 PRECURSOR.
GN          THBS1 OR TSPL OR TSP-1.
OS          Bos taurus (Bovine).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae.
OC          Bovidae; Bovinae; Bos.
OX          NCBI_TaxID=9913;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=HOLSTEIN; TISSUE=Tooth.
RX          MEDLINE=98173773; PubMed=3507054.
RA          Ueno A., Yamashita K., Nagata T., Tsurumi C., Miwa Y., Kitamura S.,
RA          Inoue H.;
RT          "cDNA cloning of bovine thrombospondin 1 and its expression in
RT          odontoblasts and predentin."
RT          Biochim. Biophys. Acta 1382:17-22(1998).
RN          [2]
RP          SEQUENCE OF 1-18 AND 710-1170 FROM N.A.
RC          TISSUE=Aortic endothelium;
RA          Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RA          Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC          -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC          CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN
CC          LAMININ AND TYPE V COLLAGEN. MAY PLAY A ROLE IN DENTINOGENESIS
CC          AND/OR MAINTENANCE OF DENTIN AND DENTAL PULP.
CC          -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
CC          -1- TISSUE SPECIFICITY: ODONTOBLASTS.
CC          -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.

```


CC	-1- SIMILARITY: CONTAINS 1 WFCC DOMAIN.	
CC	-1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.	
CC	-1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.	
CC	-1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AB005287; BAA21115.1; -	
DR	EMBL; X87618; CAA60950.1; -	
DR	EMBL; X87619; CAA60951.1; -	
DR	HSSP; P35555; IEMO	
DR	GlycoSuiteDB; Q28178; -	
DR	InterPro; IPRO00561; -	
DR	InterPro; IPRO00884; -	
DR	InterPro; IPRO01007; -	
DR	InterPro; IPRO01881; -	
DR	Pfam; PF00008; EGF; 2.	
DR	Pfam; PF00090; tsp_1; 3.	
DR	Pfam; PF00093; vwc; 1.	
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.	
DR	PROSITE; PS01186; EGF_2; FALSE_NEG.	
DR	PROSITE; PS01208; WFCC; 1.	
DR	PROSITE; PS50092; TSP1; 3.	
KW	Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;	
KW	EGF-like domain; signal.	
FT	SIGNAL	1
FT	CHAIN	19
FT	DOMAIN	19 1170
FT	DOMAIN	19 232
FT	DOMAIN	316 373
FT	DOMAIN	379 430
FT	DOMAIN	435 491
FT	DOMAIN	492 548
FT	DOMAIN	549 587
FT	DOMAIN	588 645
FT	DOMAIN	646 690
FT	DOMAIN	723 758
FT	DOMAIN	759 781
FT	DOMAIN	782 817
FT	DOMAIN	818 840
FT	DOMAIN	841 878
FT	DOMAIN	879 914
FT	DOMAIN	915 950
FT	DOMAIN	951 1170
FT	SITE	926 928
FT	DISULFID	270 270
FT	DISULFID	274 274
FT	DISULFID	551 562
FT	DISULFID	556 572
FT	DISULFID	575 586
FT	DISULFID	592 608
FT	DISULFID	617 617
FT	DISULFID	620 644
FT	DISULFID	650 663
FT	DISULFID	657 676
FT	DISULFID	678 689
FT	CARBOHYD	248 248
FT	CARBOHYD	360 360
FT	CARBOHYD	708 708
FT	CARBOHYD	1067 1067
FT	CARBOHYD	1085 1085
FT	CONFLICT	805 805
Q	SEQUENCE	1170 AA: 129533 MW: 5006ADGF3E5FA031A CRC64:

Query Match	2.6%;	Score 166;	DB 1;	Length 1170;
Best Local Similarity	20.2%;	Pred. No. 0.0012;		
Matches 109;	Conservative 41;	Mismatches 161;	Indels 228;	Gaps 26

QY	10	SEFLVETATSVLGRFMTLTGTVRRVGKKMPSTIFAPADTLSP-----GEW	56
Db	278	SSMWLELRGLRTIYTLTQDSIRKYTEBK-----ELADELRPRPLCTHNGVYRTGDW	331
QY	57	TT-----LFNIDPGKGKDYEERLDIAFREYYGGDRCARPRL	92
Db	332	TVDSCTECRCONSNTYLTKKVSCEIPMCSNATVPDGE-----CCRPWPS	375
QY	93	EARTTDMPAGSTGVVHGSPREGFW-----CLNEQRPGONGSNITYRFLCPFGSLR	146
Db	376	DSADDGWSPWSE-----WTSGSVTCGNISIQGRGSCDSLNNR--CEGSSVQT	420
QY	147	DTERI-----WSPSPSKSAACGGTGVTPTRICLAEMVSL-----CSEASE	190
Db	421	RTCHIQPCDKRFKODGSHMSWSPSSCSYSTGCD-GVITTRICLNSPSPQMCKPECGKAR	479
QY	191	EGQHGMQDC-----TACDLTC-----PMGOVNA-DC-----D	217
Db	480	ETFKACQDCSDPINCWGWPMSPMDICSVTCGGGVOKRSRLCNPNPKFOFGKDDGVDTENQ	539
QY	218	ACMGQDMLGAVSLPEGASASAALTYLLTKPKILLTQDSDRGRRIGQL----DGK	272
Db	540	ICKNDQPDIDGCJLSNPCEAVQ-----CTSYF-----DGSWK-C-GACPFGYSGGDV	584
QY	273	SILKITVFKEAPIYVLTMPKTSLKAAATIKAEFYAETPPYWMVMEPTKARAGOSVSLCCA	332
Db	585	ECKNVDECK-----EYPDACFHNHNEHCENTDPGYNC--	617
QY	333	TGKRPRPKTYWYNHTDLLDPESLYKHESKLVLKRLQHOAGEFYCKAOSDGAVKVAQL	392
Db	618	-LDCPPRF-----TGSOPGRCRVEHATANKVCVK-----	645
QY	393	IIVTASDETPCNPVESYLRLPHDFCONFTNSP---YYD-VGRCPVTKCAQQDONGIRC	447
Db	646	-----PRNPCTDG-----THDCKNKMAKNILGHYSDPMTRCCK-PGIAGNGITC	689

RESULT 4
 TSP2_BOVIN ID TSP2_BOVIN STANDARD; PRT: 1170 AA.

AC 095116; 028180;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 2 PRECURSOR (CORTICOTROPIN-INDUCED SECRETED PROTEIN) (CISP).
 OS Bos taurus (Bovine).
 OC Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 NCBI_Taxid=9913;
 RP SEQUENCE FROM N.A.
 RA Danik M., Chin A., Lafeuillade M., Keramidas M., Aguesse-Germon S., Penhat A., Chen H., Mosher D., Chambaz E.M., Feige J.J.; Submitted (MAR-1998) to the EMBL/genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF I-522 FROM N.A.
 RX MEDLINE=96311130; PubMed=8698834;
 RA Lafeuillade B., Pelletier S., Keramidas M., Danik M., Chambaz E.M., Feige J.J.;
 RT "Opposite regulation of thrombospondin-1 and corticotropin-induced secreted protein/thrombospondin-2 expression by adrenocorticotrophic hormone in adrenocortical cells.";
 RL J. Cell. Physiol. 167:164-172(1996).
 RN [3]

RP SEQUENCE OF 318-831 FROM N.A.
 RC TISSUE=Aortic endothelium;
 RA Zafar R.S., Moll Y.D., Momack J.F., Walz D.A.;
 RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of TGF-beta.";

RA Crocker P.R.;
 RT "Purification and properties of stialoadhesin, a stialic acid-binding
 RL receptor of murine tissue macrophages.";
 RL EMBO J. 10:1661-1669(1991).
 [4]
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 20-138.
 RA MEDLINE-98325385; PubMed-9660955;
 RA May A.P., Robinson R.C., Vinson M., Crocker P.R., Jones E.Y.;
 RT "Crystal structure of the N-terminal domain of stialoadhesin in
 RT complex with 3' sialylactose at 1.85 A resolution.";
 RL Mol. Cell 1:719-728(1998).
 CC -1- FUNCTION: MACROPHAGE-RESTRICTED ADHESION MOLECULE THAT MEDIATES
 CC SIALIC-ACID DEPENDENT BINDING TO CELLS.
 CC -1- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN (ISOFORM 1) AND
 CC SOLUBLE (ISOFORMS 2 AND 3).
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN SPLEEN AND LYMPH NODE
 CC WITH LOWER AMOUNTS IN LUNG, LIVER, BONE MARROW, HEART AND SKIN. NO
 CC EXPRESSION IN THYMUS, KIDNEY, BRAIN OR SMALL INTESTINE.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 16 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC
 DR EMBL: 236293; CA85290.1; -;
 DR EMBL: 236293; CA85268.1; -;
 DR EMBL: 236294; CA85269.1; -;
 DR EMBL: U92842; AAB95641.1; -;
 DR EMBL: U92843; AAB95641.1; JOINED.
 DR EMBL: U92834; AAB95641.1; JOINED.
 DR EMBL: U92836; AAB95641.1; JOINED.
 DR EMBL: U92837; AAB95641.1; JOINED.
 DR EMBL: U92838; AAB95641.1; JOINED.
 DR EMBL: U92839; AAB95641.1; JOINED.
 DR EMBL: U92840; AAB95641.1; JOINED.
 DR EMBL: U92841; AAB95641.1; JOINED.
 DR PDB: 1OP0; 16-APR-99.
 DR PDB: 1QFP; 16-APR-99.
 DR MGD: MGI:99668; Sn.
 DR InterPro: IPR003006; -;
 DR Pfam: PF00047; ig_17;
 DR PROSITE: PS00290; IG_MHC.1.
 KW Signal; Alternative splicing; Glycoprotein; Cell adhesion;
 KW Immunoglobulin domain; Transmembrane; 3D-structure.
 FT CHAIN 1 1694 SIGNAL.
 FT DOMAIN 20 1638 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1639 1659 POTENTIAL.
 FT DOMAIN 1660 1694 CITOPASMIC (POTENTIAL).
 FT DOMAIN 20 136 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 153 225 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 256 313 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 340 398 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 427 499 IG-LIKE C2-TYPE DOMAIN 4.
 FT DOMAIN 525 593 IG-LIKE C2-TYPE DOMAIN 5.
 FT DOMAIN 618 692 IG-LIKE C2-TYPE DOMAIN 6.
 FT DOMAIN 718 777 IG-LIKE C2-TYPE DOMAIN 7.
 FT DOMAIN 806 879 IG-LIKE C2-TYPE DOMAIN 8.
 FT DOMAIN 904 962 IG-LIKE C2-TYPE DOMAIN 9.
 FT DOMAIN 1069 1093 IG-LIKE C2-TYPE DOMAIN 10.
 FT DOMAIN 1095 1151 IG-LIKE C2-TYPE DOMAIN 11.
 FT DOMAIN 1181 1243 IG-LIKE C2-TYPE DOMAIN 12.
 FT DOMAIN 1269 1326 IG-LIKE C2-TYPE DOMAIN 13.
 FT DOMAIN 1355 1428 IG-LIKE C2-TYPE DOMAIN 14.
 FT DOMAIN 1455 1515 IG-LIKE C2-TYPE DOMAIN 15.
 FT DOMAIN 1544 1617 IG-LIKE C2-TYPE DOMAIN 16.

FT	SITE	827	829	CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID	41	98	BY SIMILARITY.
FT	DISULFID	160	218	BY SIMILARITY.
FT	DISULFID	263	306	BY SIMILARITY.
FT	DISULFID	347	391	BY SIMILARITY.
FT	DISULFID	434	492	BY SIMILARITY.
FT	DISULFID	532	576	BY SIMILARITY.
FT	DISULFID	625	685	BY SIMILARITY.
FT	DISULFID	725	770	BY SIMILARITY.
FT	DISULFID	813	872	BY SIMILARITY.
FT	DISULFID	911	955	BY SIMILARITY.
FT	DISULFID	1000	1062	BY SIMILARITY.
FT	DISULFID	1102	1144	BY SIMILARITY.
FT	DISULFID	1188	1236	BY SIMILARITY.
FT	DISULFID	1276	1319	BY SIMILARITY.
FT	DISULFID	1362	1421	BY SIMILARITY.
FT	DISULFID	1462	1508	BY SIMILARITY.
FT	DISULFID	1551	1610	BY SIMILARITY.
FT	CARBOHYD	159	159	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	266	266	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	299	299	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	340	340	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	500	500	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	583	583	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	693	693	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	722	722	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	737	737	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	882	882	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1089	1089	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1099	1099	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1246	1246	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1459	1459	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1473	1473	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	326	340	MAEVMKMPAGPYLEN -> SESMRLRGVSGKH (IN ISOFORM 2).
FT	VARSPLIC	341	1694	MISSING (IN ISOFORM 2).
FT	VARSPLIC	1528	1598	YTPKPTPLIVEPQGGHGGIDCDYDSEPLAILLHRSQ LVAHQHDAKPKPHRYTAPNALRVDE -> CEYEPIS ALCSLHGTGPQYQASSAOSKFGIGKGLTLASLACWME VSMLGYPALKMILLPDMFEYR (IN ISOFORM 3).
FT	VARSPLIC	1599	1694	P -> Q (IN AA SEQUENCE).
FT	CONFLICT	590	590	IHF -> FLV (IN AA SEQUENCE).
FT	CONFLICT	1049	1051	LE -> VQ (IN AA SEQUENCE).
FT	CONFLICT	1054	1055	T -> Q (IN AA SEQUENCE).
FT	CONFLICT	1061	1061	S -> Q (IN AA SEQUENCE).
FT	CONFLICT	1065	1065	
SO	SEQUENCE	1694 AA; 183087 MW; 1D2BFB39B7299C8C CRC64;		

Query Match 2.5%; Score 161; DB 1; Length 1694;
 Best Local Similarity 20.9%; Pred. No. 0.0047;
 Matches 130; Conservative 71; Mismatches 224; Indels 196; Gaps 26;

QY	8	VSEFVLEVTISVIGROTMLTQSVRRVQPGKKN-----PSIFAKADPLESGEGETITLFNID 63
DB	4	LEFSLLLASVYSGGQTTWGVSSPKNVQGLSGSCLLPICFISPADVPSNG-ITAIWYVD 62
QY	64	YPSGK-----GDYERLDAIRFYGDRCVRCARPLRLPARTDWTDPAGSGOVVHSGREGF 117
DB	63	YSKROVYVHSGPKLVDK-----RFRGR-----AELMGMDHKV----- 97
QY	118	WC--LNRQRPQGCNSNYVRF-----LCPSGLRRDTE 149
DB	98	CLLLKDLKP-EDSGTYNFRFEISDSNRMLDYKGTVTYTTDPSPTITIPBELREGME 155
QY	150	RWSPWSPWSPKSCSAACGGQGVQTRTRICLAENVSLCSSESEBQCHMGODCTACADITCGM 209
DB	156	RNFNCSTPY-----LCQEKQVSLQWNGDPTHTSVTSFOSLEPTGCV 198
QY	210	GVNNAQCQMODFMLHGVSL---PGAPASGAAYL-LIKTPK-----LLTQTPSDGR 261
DB	199	HQTTLH-----MALSWDHGKTLICQPSLGAHSSKREYVYLQVPAKGVETLLS---SSGR 251

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M25631; AAA36741.1; -
 DR EMBL; X04665; CAA28370.1; -
 DR EMBL; X14787; CAA32889.1; -
 DR EMBL; J04835; AAA61178.1; -
 DR EMBL; M99425; AAB59366.1; -
 DR PIR; A05172; A05172.
 DR PIR; A25812; A25812.
 DR PIR; A26155; A26155.
 DR PIR; A30140; A30140.
 DR PIR; A34274; A34274.
 DR HSSP; P35555; IEMO.
 DR MIM; 188060; -
 DR InterPro; IPRO00561; -
 DR InterPro; IPRO00884; -
 DR InterPro; IPRO01007; -
 DR Pfam; PF00008; EGF_1; 3.
 DR Pfam; PF00090; EGF_1; 3.
 DR Pfam; PF00093; VWC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; VWC; 1.
 DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat; EGF-like domain; Signal.
 KW EGF-like domain; Signal.
 FT CHAIN 1 18
 FT SIGNAL 1 18
 FT DOMAIN 19 1170
 FT DOMAIN 19 232
 FT DOMAIN 316 373
 FT DOMAIN 379 430
 FT DOMAIN 435 491
 FT DOMAIN 492 548
 FT DOMAIN 549 587
 FT DOMAIN 588 645
 FT DOMAIN 646 690
 FT DOMAIN 723 788
 FT DOMAIN 759 781
 FT DOMAIN 782 817
 FT DOMAIN 818 840
 FT DOMAIN 841 878
 FT DOMAIN 879 914
 FT DOMAIN 915 950
 FT DOMAIN 951 1170
 FT SITE 926 928
 FT DISULFID 270 270
 FT DISULFID 274 270
 FT DISULFID 551 552
 FT DISULFID 556 572
 FT DISULFID 575 586
 FT DISULFID 592 608
 FT DISULFID 599 617
 FT DISULFID 620 644
 FT DISULFID 650 663
 FT DISULFID 657 676
 FT DISULFID 678 689
 FT CARBOHYD 248 248
 FT CARBOHYD 360 360
 FT CARBOHYD 708 708
 FT CARBOHYD 1067 1067
 FT CONFLICT 84 84
 FT CONFLICT 523 523
 FT SEQUENCE 1170 AA; 129412 MW; 69B3DE5AE3A395E CRC64;

Query Match 2.5%; Score 157; DB 1; Length 1170;
 Best Local Similarity 20.1%; Pred. No. 0.0053;
 Matches 111; Conservative 37; Mismatches 150; Indels 254; Gaps 27;
 10 SEVLVETVGLGRQTMLTQSVRRVQPKRNPSTFAPADPLESP-----GEM 56
 278 SSMVLELRGLRTIVTTLQDSIRKVTENK-----ELANELRRPPLCYHNGVYRNNEEM 331

QY 57 TT-----LFNIDYPGKGDYERLDAIRFYGGDRVCARPLRL 92
 Db 332 TVDSCTFCHQNSVTICKKVSCTPIMCNSNATVPDGE-----CCPCPMWS 375
 QY 93 EARTTWTPTAGSTGQYVHGSPREGFW-----CLNRQRRGQNSYTVAFELCPSSLRR 146
 Db 376 DSADDGWSPWSE-----WTSCSTSCGNSIQGRNSCDLNNN--CEGSSVOT 420
 QY 147 DTERI-----WSPWSPMSKCSACGOTGYOTRTRICLAWVSL--CSEASE 190
 Db 421 RTCHIDECCKRFKQDGGWSHMSWSSCSVTCCD-GVITTRICLNSPSPQWNGKPCGEGAR 479
 QY 191 EGQHCWGQDC-----TACDLTC-----PMGVNA-DC-----D 217
 Db 480 ETRACKKDACPTINGMGWPMSPMDICSVTCGGGYKRSRLCNNTPPGCGKDCVGVTENG 539
 QY 218 ACMQDPEMLHGANVSLGGAPASGAATYLLTKTKLLTQDSDGRFRIPGLCPGKSLIKI 277
 Db 540 ICKKQDCPTIDGCLSNPCFAVR-----CTSP-----DSWKC-GACPPGYS----- 580
 QY 278 TKVFPVLTMPKTSLKATIAEFVRAETPYWVNNPETKARAGQVSLCKATGKPR 337
 Db 581 -----GNGIQCTDVDECKEY 595
 QY 338 PDKRYFVHNDTLDPSTLYKHESKLVLRKLOHQAGPYFCKAQSADG----- 383
 Db 596 PDACF-NHN-----GEHRE-MDPGYNCLPCPPRFTGSG 628
 QY 384 ----AKSKYAQOLIYASDETPCNPYPESILNLPDCCQONATNSF---YTD-VGRCPPK 435
 Db 629 PFGQVEHATANKQVC---KPRNPCTDG---THDCNKNACNYLGHSDPMYRCECK 679
 QY 436 TCAGQDNGIRG 447
 Db 680 --PGYAGNGLIC 689

RESULT 9
 PGHM_HUMAN STANDARD; PRT; 4393 AA.
 ID P98160; O16287;
 AC P98160; O16287; (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 01-OCT-1996 (rel. 34, Last annotation update)
 DE BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
 DE PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
 GN HSPG2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=92112994; PubMed=1730768;
 RA Kallunki P., Tryggvason K.;
 RT "Human basement membrane heparan sulfate proteoglycan core protein: a
 RT 467-kD protein containing multiple domains resembling elements of the
 RT low density lipoprotein receptor, laminin, neural cell adhesion
 RT molecules, and epidermal growth factor."
 RL J. Cell Biol. 116:555-571(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin, and Colon;
 RC MEDLINE=92235084; PubMed=1569102;
 RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
 RT "Primary structure of the human heparan sulfate proteoglycan from
 RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
 RT domains homologous to the low density lipoprotein receptor, laminin,
 RT neural cell adhesion molecules, and epidermal growth factor."
 RL J. Biol. Chem. 267:8544-8557(1992).
 RN [3]
 RP SEQUENCE OF 1018-1472 FROM N.A.


```

FT SITE 65 67 HEPARAN SULFATE (POTENTIAL).
FT SITE 71 73 HEPARAN SULFATE (POTENTIAL).
FT SITE 76 78 HEPARAN SULFATE (POTENTIAL).
FT SITE 4151 4153 MEDIATES MOTOR NEURON ATTACHMENT
(POTENTIAL).
FT SITE 4301 4303 MEDIATES MOTOR NEURON ATTACHMENT
(POTENTIAL).
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 234 BY SIMILARITY.
FT DISULFID 285 297 BY SIMILARITY.
FT DISULFID 292 310 BY SIMILARITY.
FT DISULFID 304 319 BY SIMILARITY.
FT DISULFID 325 337 BY SIMILARITY.
FT DISULFID 332 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 375 394 BY SIMILARITY.
FT DISULFID 388 403 BY SIMILARITY.
FT DISULFID 766 775 BY SIMILARITY.
FT DISULFID 768 782 BY SIMILARITY.

```

Query Match Best Local Similarity 2.4%; Score 151.5; DB 1; Length 4393;

Matches 149; Conservative 84; Mismatches 263; Indels 233; Gaps 36;

```

OY 84 RVCARPLRLARTTWTTPAGSTGVVHSGSPREG-FWCINREORPGONCSNTYVRELCPP 141
DB 3150 RISSTPAKLEORTYGLMDSHYVLQISSAKPSDAGTYVCL-----AQNL----- 3193
OY 142 GSIRRTERT-----WSPSPKSCSA-----CGOTGVQVRT----- 174
DB 3194 GAAQKQVEYIVDTGAMAPAPPOVAEAEIYVAGHTATLRCSATSPARTIMSKLRSP 3253
OY 175 -----RICLAEMYSLCSEASESGHMGODCTACDLTCPMGVNA----- 214
DB 3254 LPMQHLEBDTLIRVAAODSGOY-----ICNATSPAGHAETIILHVESPYATYV 3306
OY 215 -----DCQACMCODFMHSG-----AVSLPGADASGAAYLLTKTKPLLT 254
DB 3307 PEHASVAGEETVLOLC-----LAHGTPLPTFOWMSVSGSLPGRATARNELHFERAP----- 3359
OY 255 QVDSQGRFRTPLGLCPDCKSLIKIT-KV-----KRAPIVLIMPKSLKAATIKAEVNAETP 309
DB 3360 --EDSGRYR-----CRVTNKGSAEAPQOLVGGPESLPATSIPIA-----GSP 3402
OY 310 YVMWNETARRAGOSVSLCKKATGKRPDKYFWYHNDLLDPSLKHESKIVLRKLOOH 369
DB 3403 TVQVTPQLTKTSIGASVEFHC-AVPSDRGTQLRMFKEGQLPRGHSVQGVLRIONLDOS 3461
OY 370 QAGEYFCKAQSDAGAVKSKVAQOLIVTASDETCNPYESTILRLPHDCEONATNSFYDV 429
DB 3462 CGGTTCQAHHGPMGKAQAS-AQLVIA-----LP-SVLINI----- 3495
OY 430 GRCPVATCAGQDNGIRCDAYONCCGISKTERETIOCGSYTLPTVAKECSQRTETR 489
DB 3496 -----RTSVQTVV-VGHAVEECLALGDKPR-----QVTW 3524
OY 490 SIYRG-----RVSADNGEPMRFEGHYM-----GNSRYSMTGKGTFTLHVPODTERLVLFVVR 544
DB 3525 SKVGHGLRGIYSGGVVRIAHVELADACQYRTATNAGT-----TQSHVLLVQA 3576
OY 545 LQKVFVTVVLPNNKGSVAFHEIKMLRRKEPITLLEAMETNIIPLGEVNGE-DPMAELE- 602
DB 3577 LPQ-ISMPEVFRVAPGSAAVE-----PCIASGYPTLPDISWSKLDGSLPRDRLEN 3625
OY 603 -----ISRSFYRQNGEPIY-----GKYKASTFLDPRNISTATAQDIDLANINDEGDT 651
DB 3626 NMLMLP--SVQPDADAGTYCTATNRQGVKAKAFANLQVPERV-VPYPTQTPYSEFL----- 3676
OY 652 FLPLRTGMSVDFRDEVTSEPLNAGVKVYKHLSDTQVKMEHISTVYKLSLNP-----TGLM 708
DB 3677 -PLPTIKDAVRRKEIKITRPRPSADGMLX--NGQKRVG--SPYMLANRQDPFISFGLV 3731

```

```

OY 709 EEEGDKEFE 717
DB 3732 GGRPERFRD 3740

RESULT 10
NRG.DROME STANDARD; PRT: 1239 AA.
ID P20241: Q24414.
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-2000 (Rel. 10, Last annotation update)
DE NEUROGLITAN PRECURSOR.
GN NRG.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta.
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-41 AND 737-751.
RX MEDLINE=90030418; PubMed=2805067;
RA Bieber A.J., Snow P.M., Hortsch M., Patel N.H., Jacobs J.R.,
RA Traquina Z.R., Schilling J., Goodman C.S.;
RT "Drosophila neuroglian: a member of the immunoglobulin superfamily
RT with extensive homology to the vertebrate neural adhesion molecule
RT L1."
RL Cell 59:447-460(1989).
RN [2]
RP SEQUENCE OF 1182-1239 FROM N.A.
RX MEDLINE=90262720; PubMed=1693086;
RA Hortsch M., Bieber A.J., Patel N.H., Goodman C.S.;
RT "Differential splicing generates a nervous system-specific form of
RT Drosophila neuroglian."
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 610-814.
RX MEDLINE=94213741; PubMed=7512815;
RA Huber A.H., Wang Y.-M.E., Bieber A.J., Bjorkman P.J.;
RT "Crystal structure of tandem type III fibronectin domains from
RT Drosophila neuroglian at 2.0 A."
RL Neuron 12:717-731(1994).
CC -1- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN NEURAL AND GLIAL CELL
CC -1- ADHESION IN THE DEVELOPING DROSOPHILA EMBryo.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC -1- SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: NEURONS AND GLIA IN THE DEVELOPING NERVOUS
CC -1- SYSTEM AND ON SOME OTHER NONNEURONAL TISSUES.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: M28231; AAA28728.1; ALT_SEQ.
DR EMBL: X76243; CA53822.1; -.
DR PIR: A32579; A32579.
DR PDB: 1CFB; 30-NOV-94.
DR FlyBase: FBgn0002868; Nrg.
DR InterPro: IPR001777; -.
DR InterPro: IPR003006; -.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; lg; 6.
DR Cell adhesion: Glycoprotein; Transmembrane; Repeat; 3D-structure;
KW Immunoglobulin domain; Signal; Embryo; Alternative splicing.
FT SIGNAL 1 23

```

```

FT CHAIN 24 1239 NEUROGLIAN.
FT DOMAIN 24 1138 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 1139 1154 POTENTIAL.
FT DOMAIN 1155 1239 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 53 123 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 149 124 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 262 329 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 354 422 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 447 512 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 536 606 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 629 690 FIBRONECTIN TYPE-III.
FT DOMAIN 729 792 FIBRONECTIN TYPE-III.
FT DOMAIN 832 896 FIBRONECTIN TYPE-III.
FT DOMAIN 932 997 FIBRONECTIN TYPE-III.
FT DOMAIN 1024 1098 FIBRONECTIN TYPE-III.
FT DISULFID 59 111 POTENTIAL.
FT DISULFID 625 706 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 683 683 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 821 821 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1234 1234 T -> Y (IN REF. 2).
FT CONFLICT 1237 1237 L -> K (IN REF. 2).
SQ SEQUENCE 1239 AA; 138284 MM; 49EL2692DDDD027D CRC64;

```

Query Match Best Local Similarity 2.3%; Score 146; DB 1; Length 1239; Matches 149; Conservative 99; Mismatches 277; Indels 208; Gaps 41;

```

QY 304 VRAETPYVMNPEFKARRAGOSVSLCKATGKPRPKXYFWYHNDLDPDLKHEK--L 361
D 427 VQAEPTISEAPAVSVTDGRNVTIKCRVNGSPKP-LVKMLRASWMLGGRYVNGDGL 485
QY 362 VLRLQOHQAGEYFCKKASDAGAVSKVAQLIVASDETPCNPPPESTLILPHDCFQNA 421
D 486 EIDQVTFSDACKYCYAONKPFGEIAD-GSLIV--KEHTRITOEPONEYEA----AGOSA 538
QY 422 T---NSFYDVGRCPRVTKCA--GO-----ODNGIRCRRAVONCGISKEEREIOCSGYT 471
D 539 TFRONEAHDLLEIEMDKGOSIDEPAQRFVKYND--NSLTIAIT--MELDSGEYT 593
QY 472 LPTVARECSQCRCTEERSIVR-----GRVSAADNGEPMRGHYMGNRSYMTGY 522
D 594 C---VARTRLDEATARANLIVQDVNAPKLTGICQADKAE---IHWEOQGDNSPLIHY 647
QY 523 KGTG-TLHVP-----QDTERLVLTFFYDLQKTFVN--TTKVLPRNKKGSA--VHEIKML 571
D 648 TIOENTSTFPASMDAAVEKVENPTDSSFVQMSPMANFYFVIAFNKIGASPPSAHSOSCT 707
QY 572 RR-----KEPTLEAMETNIIPLGEVVGEDMAELE-----IPSRNF--- 608
D 708 TQDPVPRKNPNVVGQGTG--PNNLVISWTMPLEIENHAPNFHYVSKRILPAAMENN 765
QY 609 ---YRONG---EPYIGKVASVTFDPNPNISTATTAQOITDLNFINDEGDTFPLRTYGM 659
D 766 NIFPMRONNIYIADQPTFVKLIKVAINDRGESVNAEEV---VQSGSDRPLDAPTIN 821
QY 660 FSVSPRDEVTSE-----PLNAGKAVHLDSIOVKMPEHISTYVKLMSLNDPTGLMEEE 711
D 822 FTM--ROITSTSGYAMTPIVSESVGHFKGYK-----IQWTV-----ENE 861
QY 712 GDFPFENOR--RNKREDPTFLVGNLEIRERLJNDVPESRCFCAVAVSERFLPSEQ 769
D 862 GE---EGIRELHVAKDTHNALVTOFK-----PDSKN-YAKILAYIKR----- 899
QY 770 IQGVVIVINLEPRTGLSNPRANGRFDPSVITGPNGACVPAFCDDQSDPAY---SAYVL 825
D 900 -----FNGPPSAVIDFDT-----PEGVSPV-----QGLDAVPLIGSASATML 935

```

```

QY 826 -----ASTAGEELOAVESPKEFNMAIGVPOPLYLNRYRRTHDEPRVKKTAFOI 876
D 936 HMKRPLPNGLTGTGKIYEE-----VASEYIGERREDPHITDPRVTR-----M 980
QY 877 SMARPPNS-----AEESNGPIYAFE--NLRACEAPPSAHPFYQIECD----R 921
D 981 KMGKGLPKSKYRISITATTKGEGSEHYIEKTLTKDAVNVAAPVPPSWEGQLPSDNGLAK 1040
QY 922 YDVTVPFNEDDP 934
D 1041 FRIMWLPSTEGHP 1053
RESULT 11
TSPL_XENLA STANDARD; PRT; 1173 AA.
ID TSPL_XENLA AC P35448;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 1 PRECURSOR.
GN THBS1 OR TSPL.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Urry L.A., Ramos J., Duquette M., Desimone D.W., Lavrier J.;
RL Submitted (xxx-1993) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L04276; -; NOT_ANNOTATED_CDS.
DR HSSP; P00740; IIXA.
DR InterPro; IPR000561; -.
DR InterPro; IPR000884; -.
DR InterPro; IPR001007; -.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF00093; vwc; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSPL; 3.
DR PROSITE; PS01208; vwc; 1.
DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1173 THROMBOSPONDIN 1.
FT DOMAIN 23 235 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 319 376 VWFC.
FT DOMAIN 382 433 TSP TYPE-1 1.
FT DOMAIN 438 494 TSP TYPE-1 2.
FT DOMAIN 495 546 TSP TYPE-1 3.
FT DOMAIN 550 590 EGF-LIKE 1.
FT DOMAIN 591 648 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

```


DR EMBL: M96655; AAA49083.1; -
DR EMBL: M88283; AAA8647.1; -
DR EMBL: M88284; AAB53768.1; -
DR PIR: A25810; A25810.
DR PIR: A37099; A37099.
DR PIR: S11652; S11652.
DR InterPro: IPR000719; -
DR InterPro: IPR001777; -
DR InterPro: IPR002290; -
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00047; Ig; 9.
DR Pfam: PF00069; Pkinase; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_SF; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW transglutaminase; Serine/threonine-protein kinase; Calmodulin-binding;
KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
KW Alternative initiation.
FT CHAIN 1 1906
FT CHAIN 935 1906
FT CHAIN 1750 1906
FT INIT_MET 935 1906
FT INIT_MET 1750 1906
FT DOMAIN 1330 1400
FT DOMAIN 1453 1708
FT DOMAIN 1809 1876
FT NP_BIND 1459 1467
FT BINDING 1482 1482
FT AC_SITE 1574 1574
FT DOMAIN 1716 1728
FT DOMAIN 1730 1749
FT DOMAIN 1317 1364
FT DOMAIN 1385 1402
FT DOMAIN 660 1833
FT REPEAT 660 676
FT REPEAT 758 774
FT REPEAT 1107 1123
FT REPEAT 1817 1833
FT DOMAIN 693 1866
FT REPEAT 791 807
FT REPEAT 1140 1156
FT REPEAT 1281 1297
FT REPEAT 1851 1866
FT DOMAIN 970 1226
FT REPEAT 970 987
FT REPEAT 999 1016
FT REPEAT 1061 1078
FT REPEAT 1209 1226
FT DOMAIN 1700 1763
FT DOMAIN 1896 1906
FT MOD_RES 1748 1748
FT MOD_RES 1762 1762
FT CONFLICT 1439 1439
SQ SEQUENCE 1906 AA; 210445 MW; AD7DBAJB69EE363 CRC64;

Query Match 2.38; Score 145; DB 1; Length 1906;
Best Local Similarity 19.38; Pred. NO. 0.074;
Matches 179; Conservative 109; Mismatches 305; Indels 334; Gaps 45;

QY 140 PGSLRPTERISWSPWSKCSAAGQGTGTRICIAEVLSEASEE-----GQ 193
DB 1111 PPAVY-----SWTLDSK-----AIKSSKSIYISQEGTICSLTIKVMEDDGE 1153

QY 194 H-CMGQDTACDLTCGPMQVNAADCAACQCPMLHGAIVSLFGAPASCAAIYLLTKTKPL 252
DB 1154 XCIAENAG-----KAEC-AC---RVLVEDTSSTKAAKPAEKTKPKTTLPPV 1199

QY 253 LTQTDSDGRFRLPGICPDGKSLIKTKYKPAPIVLTMPKTSIAKAEVRAETPYMV 312
DB 1200 LTESSE-----ATVKKRP-----AKTPPKAT-----PQIT 1228

QY 313 MNPETKARAGOSVSLCCATGKPRPDYFYWHNDTLDPSTLY-----KHESKLVLRKL 366
DB 1229 QPEDRKVRAGESVLFPAKVVG--TPRTCTMVKFRKQIQEENYIKIENAESSKLTIST 1287

QY 367 QOHQAGEYFCKAQSDAGAKSKVAQLIVTASDE--TPC-NPVPSYLIRLPDHFQONAT 422
DB 1288 KOEHGCGYTLVVENKLSROAQVNLTVDKDDPPAGTPCASDIHSSSL-----T 1336

QY 423 NSFY--DVCRC-----PYKTCAG-----QOONGICRCPAONC 454
DB 1337 LSWGSSYDGSANOSYVEIWNVSVDNKNWTDLTTCRSTSFNVQDLQADREYKFRRAANV 1396

QY 455 CGIS-----KTEEREIQSGYTLPTKYAKESQCRCTETRSIVGRVSAADNG 502
DB 1397 YGISPQSESEVYKVGKEQEELKEEAEISDDECKETEVNRYRTTN--TEOKVSDVYNI 1455

QY 503 EPM---RFGHYVGNNSVSMTGKGT-----TLHPDPTERLV 538
DB 1456 EERLGSFGGVFLVERKTKGVMAKFFKAYSAKEKENIRDEISIMNCLHHPK-----L 1510

QY 539 LTFVDRLOKFNVTFTVLPFNKKGSAVFHEIKMLRKKEPTITLEAMETNITPLGEVVGEDPM 598
DB 1511 VQCVDAFEKANIYVWLDL--VSGGELFERI-----IDED-- 1543

QY 599 AELEIPSR---SFYRO--NGEPYIGKRVASYTFDLPNISTATAQDTLNFINDGDFP 653
DB 1544 --FELTERECIKYMQQISEGVEYIHKQGVHLDLKPENIMCVNKTGTSIKLID----- 1594

QY 654 LRTYGVSVDFRDEYTSPLNAGKVKVHLDSYQVMPHIS-----TVKLMS----- 700
DB 1595 -----FOLARLE-----SAGSLKLVFGTPEFVAPVINEPIGETDWMISGIVCIYI 1642

QY 701 ---LNPDTGLMEER-----GDEKFNORRNK--REERTFLVGNLEIRERFLVDPE 748
DB 1643 LVSGISPPMGNDNETLANVTSATWDFDEADPDSADAKPDIS-----LKKDKMS 1695

QY 749 SRRC-----FVKVAVRSERFLPSEIQGVISVINLEPRTGFLSNPR 791
DB 1696 RLNCQCCQHHWLOKDTNMEAKKLSKDRMKKYMARRKW-----KTGNA--VR 1742

QY 792 AMGRPDS--VITGPNGACVPAFCDDQSPDAYSAVYLAELQAVSSPKFNPNAIGV 849
DB 1743 AIGRLSSAMISGMSGR-----KAGSSP----- 1766

QY 850 POPYLNTKNTYRTHEDRVRKKTAFQISMAMPKPRPSAESNGPIYAFENLACEAPPSA 909
DB 1767 ---TSPLNADKVENEDAFLEEVAEKPRVPR-----YFTKTLIDMEVEGSA 1810

QY 910 AHFRFYQTEGDYR--DYNTVVPFNMEDPM 935
DB 1811 ARFD-CKIEG--YDPDEVWYWKDDQPV 1834

RESULT 13
KMLS_BOVIN STANDARD; PRT; 1176 AA.
AC 028824:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE (CC 2.7.1.117) (MLCK)
DE [CONFAINS: TELOKIN].
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;

ID KML5_HUMAN STANDARD; PRT: 1914 AA.
AC Q15746; Q14644; Q16794; Q95796; Q95797; Q95798; Q95799; Q9UBG5;
DT 15-JUL-1998 (Rel. 36, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES
DE (EC 2.7.1.117) (MCK) [CONTAINS: TELOKIN (KINASE RELATED PROTEIN)
DE (RPP)].
GN MYLK OR MLCK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=umbilical vein endothelial cells;
RX MEDLINE=97304466; PubMed=9160829;
RA Garcia J.G.N., Lazar V., Gilbert-McClain L.I., Gallagher P.J.,
RA Verin A.D.;
RT "Myosin light chain kinase in endothelium: molecular cloning and
RT regulation.";
RL Am. J. Respir. Cell Mol. Biol. 16:489-494(1997).
RN [2]
RP REVISIONS.
RA Birukov K.G., Garcia J.G.N.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3A; 3B AND 4).
RC TISSUE=umbilical vein;
RX MEDLINE=99216419; PubMed=10198165;
RA Lazar V.L., Garcia J.G.N.;
RT "A single human myosin light chain kinase gene (MLCK; MYLK).";
RL Genomics 57:256-267(1999).
RN [4]
RP REVISIONS TO ISOFORM 2.
RA Birukov K.G., Garcia J.G.N.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 923-1914 FROM N.A.
RC TISSUE=Hipocampus;
RX MEDLINE=96121365; PubMed=8575746;
RA Potter M.-C., Cheiot E., Pekarsky Y., Gardiner K., Rossier J.,
RA Turnell W.G.;
RT "The human myosin light chain kinase (MLCK) from hippocampus:
RT cloning, sequencing, expression, and localization to 3qgen-q21.";
RL Genomics 29:562-570(1995).
RN [6]
RP SEQUENCE OF 1614-1914 FROM N.A.
RC TISSUE=Lung, and Placenta;
RX MEDLINE=20007838; PubMed=10536370;
RA Matterson D.M., Schavocky J.P., Guo L., Weiss C., Chlenski A.,
RA Shrivatsky V.P., Van Eldik L.J., Halech J.;
RT "Analysis of the kinase-related protein gene found at human chromosome
RT 3q21 in a multi-gene cluster: organization, expression, alternative
RT splicing and polymorphic marker.";
RL J. Cell. Biochem. 75:481-491(1999).
RN [7]
RP SEQUENCE OF 1456-1914 FROM N.A.
RC TISSUE=Placenta;
RA Matterson M.D.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
CC SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SHRINE
CC IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
CC FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS. CENTRAL
CC DETERMINANT IN THE DEVELOPMENT OF VASCULAR PERMEABILITY AND TISSUE
CC EDEMA FORMATION. IN THE NERVOUS SYSTEM IT HAS BEEN SHOWN TO
CC CONTROL THE GROWTH INITIATION OF ASTROCYTIC PROCESSES IN CULTURE
CC AND TO PARTICIPATE IN TRANSMITTER RELEASE AT SYNAPSES FORMED
CC BETWEEN CULTURED SYMPATHETIC GANGLION CELLS. CRITICAL PARTICIPANT
CC IN SIGNALING SEQUENCES THAT RESULT IN FIBROBLAST APOPTOSIS.
CC -1- CATALYTIC ACTIVITY: ATP + [MYOSIN LIGHT-CHAIN] = ADP + [MYOSIN
CC LGHT-CHAIN] PHOSPHATE.

CC -1- SUBUNIT: TELOKIN BINDS CALMODULIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS OF MLCK ARE PRODUCED
CC BY ALTERNATIVE INITIATION, A NON-MUSCLE FORM (WHICH IS THE LONGEST
CC FORM); A SMOOTH-MUSCLE FORM AND TELOKIN (A C-TERMINAL SECTION WITH
CC NO CATALYTIC ACTIVITY).
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; 1 (SHOWN HERE), 2, 3A,
CC 3B, 4 AND DEL-1790; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES ARE
CC EXPRESSED IN A WIDE VARIETY OF ADULT AND FETAL TISSUES AND IN
CC CULTURED ENDOTHELIUM WITH QUALITATIVE EXPRESSION APPEARING TO BE
CC NEITHER TISSUE- NOR DEVELOPMENT-SPECIFIC. NON-MUSCLE ISOFORM 2 IS
CC THE DOMINANT SPICE VARIANT EXPRESSED IN VARIOUS TISSUES. TELOKIN
CC HAS BEEN FOUND IN A WIDE VARIETY OF ADULT AND FETAL TISSUES.
CC -1- PTM: MLCK IS PROBABLY DOWN REGULATED BY PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO THE SPR/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U48959; AAC18423.2; -;
DR EMBL: AF069601; AAD15921.2; -;
DR EMBL: AF069602; AAD15922.1; -;
DR EMBL: AF069603; AAD15923.1; -;
DR EMBL: AF069604; AAD15924.1; -;
DR EMBL: X85337; CAA59685.1; -;
DR EMBL: AF096771; AAD51380.1; -;
DR EMBL: AF096766; AAD51380.1; JOINED.
DR EMBL: AF096767; AAD51380.1; JOINED.
DR EMBL: AF096768; AAD51380.1; JOINED.
DR EMBL: AF096769; AAD51380.1; JOINED.
DR EMBL: AF096770; AAD51380.1; JOINED.
DR EMBL: AF096771; AAD51381.1; -;
DR EMBL: AF096772; AAD51381.1; -;
DR EMBL: AF096769; AAD51381.1; JOINED.
DR EMBL: AF096770; AAD51381.1; JOINED.
DR EMBL: X90870; CAA62378.1; -;
DR HSSP: P56276; IRLK.
DR MIM: 600922; -;
DR InterPro: IPR000719; -;
DR InterPro: IPR001777; -;
DR InterPro: IPR002290; -;
DR InterPro: IPR003006; -;
DR Pfam: PF00041; fn3.1.
DR Pfam: PF00047; fn3.1.
DR Pfam: PF00069; kinase.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
KW Transferrase: Serine/threonine-protein kinase; Calmodulin-binding;
KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
KW Alternative initiation; Alternative splicing.
FT CHAIN 1 1914
FT MYOSIN LIGHT CHAIN KINASE, NON-MUSCLE
FT ISOZYME.
FT CHAIN 923 1914
FT MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE
FT ISOZYME.
FT CHAIN 1761 1914
FT TELOKIN.
FT INT_MET 923 923
FT FOR MYOSIN LIGHT CHAIN KINASE, SMOOTH-
FT MUSCLE ISOZYME.
FT INT_MET 1761 1761
FT FOR TELOKIN.
FT DOMAIN 1343 1413
FT FIBRONECTIN TYPE-III.
FT DOMAIN 1464 1719
FT PROTEIN KINASE.
FT DOMAIN 1711 1774
FT CALMODULIN-BINDING.
FT DOMAIN 1824 1891
FT IG-LIKE C2-TYPE DOMAIN.
FT NP_BIND 1470 1478
FT ATP (BY SIMILARITY).
FT BINDING 1493 1493
FT ATP (BY SIMILARITY).
FT ACT_SITE 1585 1585
FT BY SIMILARITY.


```

FT DOMAIN 1906 1914 POLY-GLU.
FT DOMAIN 868 998 5 X 28 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 868 895 1-1.
FT REPEAT 896 923 1-2.
FT REPEAT 924 951 1-3.
FT REPEAT 952 979 1-4.
FT REPEAT 980 998 1-5 (INCOMPLETE).
FT DOMAIN 999 1063 6 X 12 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 999 1003 11-1 (INCOMPLETE).
FT REPEAT 1004 1015 11-2.
FT REPEAT 1016 1027 11-3.
FT REPEAT 1028 1039 11-4.
FT REPEAT 1040 1051 11-5.
FT REPEAT 1052 1063 11-6.
FT VARSPLIC 437 506 VSGIPREYAFLEGGPVRROGSEIEVEYEDAGSHYLCILKA
FT FT RTRDSGTSCASNAGOVCSMTLOVER -> G (IN
FT FT ISOFORM 2 AND ISOFORM 3B).
FT VARSPLIC 1433 1439 DEVEYSD -> MKMRCQT (IN ISOFORM 3A,
FT FT ISOFORM 3B AND ISOFORM 4).
FT VARSPLIC 1473 1546 GKGQVFLVEKTRKVMAGKFKKAYSAKEKENIRQELISIM
FT FT NCLHHPKLVOCYDAFEKANIWMVLEIVSGGEL -> L
FT FT (IN ISOFORM 4).
FT VARSPLIC 1655 1705 MISSING (IN ISOFORM 3A AND ISOFORM 3B).
FT FT VARSPLIC 1790 1790 MISSING (IN ISOFORM DEL-1790).
FT FT CONFLICT 933 933 V -> M (IN REF. 5).
FT FT CONFLICT 963 963 S -> P (IN AAD15922).
FT FT CONFLICT 1022 1022 P -> A (IN REF. 5).
FT FT CONFLICT 1048 1050 KPM -> EAH (IN REF. 5).
FT FT CONFLICT 1162 1162 P -> L (IN AAD15922 AND AAD15923).
FT FT CONFLICT 1210 1210 L -> P (IN REF. 5).
FT FT CONFLICT 1280 1280 E -> D (IN AAD15922 AND AAD15923).
FT FT CONFLICT 1284 1284 M -> I (IN AAD15922, AAD15923 AND
FT FT AAD15924).
FT CONFLICT 1300 1300 A -> G (IN REF. 5).
FT FT CONFLICT 1316 1316 L -> S (IN REF. 5).
FT FT CONFLICT 1326 1326 T -> S (IN REF. 5).
FT FT CONFLICT 1478 1478 V -> C (IN REF. 5).
FT FT CONFLICT 1511 1511 S -> T (IN AAD15922 AND AAD15923).
FT FT CONFLICT 1553 1553 I -> T (IN REF. 5).
FT FT CONFLICT 1609 1609 A -> P (IN REF. 5).
FT FT CONFLICT 1639 1639 G -> R (IN REF. 5).
FT FT CONFLICT 1640 1640 GY -> D (IN AAD15922, AAD15923 AND
FT FT AAD15924).
FT CONFLICT 1648 1648 G -> R (IN REF. 5).
FT FT CONFLICT 1659 1659 LS -> P (IN REF. 5).
FT FT CONFLICT 1710 1711 CT -> LA (IN REF. 5).
FT FT CONFLICT 1897 1897 L -> H (IN AAD15922, AAD15923 AND
FT FT AAD15924).
SQ SEQUENCE 1914 AA: 210773 MW: 9295FE0279CB0033 CRC64:

```

Query Match 2.3%; Score 144.5; DB 1; Length 1914;
 Best Local Similarity 20.7%; Pred. No. 0.08;
 Matches 221; Conservative 133; Mismatches 357; Indels 357; Gaps 58;

```

Db 1184 QVTV-----DDAPASEN-----TKAPEKSRRRKSSLPVLGTESD----- 1219
Qy 277 ITAVKFAPIVLMPTKSLKATIKAEFVRAETPYVMNPETKARRAGOSVSLCKATGKP 336
Db 1220 -ATVKKRP-----APKTPPPRAAM-----PQIIQFEDQVVRAGESVELFGKVTG-T 1264
Qy 337 RPDKRFWYHNDTLLEDSLY-----KHESKVLRLQHQOAGEYFKCAOSDAGAVKASKVA 390
Db 1265 QPITCTWMMKFRKQIQSEHMKVENSNGSKLTLAARQEHCGCYTLLENKIGSRQAQVNA 1324
Qy 391 QLIIVTASDE---TPC-----NPVESYLRLPHCFONATNSFY 427
Db 1325 LTVVADPDPAGTPCASDIRSSSLTSLWYSSYDGSAAQVSYLEI-----WDSANKTKM 1379
Qy 428 DWGRCPVKTCAGQQ-----DNGIRCDAVONCCGISK--TEEREICSCGYTLPTVAKKCSO 483
Db 1380 ELATCSTSNVDDLPDHEKFKRAVAINVYGSFQSESLTTVG-----E 1426
Qy 484 RCTETRSIVGRVSAADNGEPMRFHVYMGNSRVMTGYKGFTHVPODTERLVLTFFVD 543
Db 1427 KPPEPRDEV--EVSDDDEKPE-----EVDYR-TVTI-----NTBQKVSDFYD 1465
Qy 544 ---RL--QKRVNTTKVLPFNKKSAY-----FHEIKMLRKEPTTLEAM----- 582
Db 1466 IEERLSSGKGFQVFRV--EKTRKWAGKFFKAYSAKEKENIRQELISIMNCLHHPKLVQ 1523
Qy 583 -----ETNIIPLGHWGVDPMAE-----LEIPSR---SFYRQ--NGEPYIGKAKAV 624
Db 1524 CVDAFEKANIWMVLEIVSGGELFERLIDEDFELTRECKIKYMRQISEVEYTHKQGIYH 1583
Qy 625 TFLDPNNISTANAOQDLNFINDEGTFPLRTYGMFSVDFRDEVTSEPLNACKVYKHLDS 684
Db 1584 LDKPENIMKVNKGTGRKIKID-----FGLARLLE-----NAGSLKLVFGT 1624
Qy 685 TOYKMEPHIS-----TVKLMS-----LNPDTGIMEEE-----GOFKRENR 720
Db 1625 PEFVAPEVNIWEPIGATYDMSIGVICIYILVSGLSPFGNDNLETANVTSATWDDDEE 1684
Qy 721 RNK-REDRTFLVGNLEIRERKRLFNLDVPESRRCF-----VKVARYRSERL 765
Db 1685 FDEISDADKOFINL-LKKMKMRKLDCTQ---CLQHPWLMKOTKKNMKAELKSLDRMKKYM 1740
Qy 766 PSEOIOGVAVISVINLEPRGFEFLSNPRAMGRFDSVTIGPAGACVPARCDOSPDAYSAYLV 825
Db 1741 ARRKWQ-----KTG--NAVRAIGRLSSMA-----MI 1764
Qy 826 ASLAGEELAVESSPEKNPAIGVPOPLYNKLNYRTHDEDDPVKKTATQISMARRPMS 885
Db 1765 SGLSGRK-----STGSP---TSPLEAKLESEED--VSOAFLFAVAEERKPHV 1807
Qy 886 AEESSNGPIYAFENIRACEAPPSAAHFRFYQIEGDRY-DYNTVPFNED 932
Db 1808 KP-----YFSKITRDLVEYEGSAAARFD-CKTIG--YPPDEVVWFRDD 1846

```

RESULT 15
 TSP2, HUMAN STANDARD; PRT: 1172 AA.
 AC P35442;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 2 PRECURSOR.
 GN THB2 OR TSP2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=94010892; PubMed=8406456;
 RA Labell T.L., Byers P.H.;
 RT "Sequence and characterization of the complete human thrombospondin 2

RT cDNA: potential regulatory role for the 3' untranslated region.";
RL Genomics 17:225-229(1993).
[2]
RP SEQUENCE OF 560-1172 FROM N.A.
RC TISSUE-Fibroblast;
RA MEDLINE=92217961; PubMed=1559694;
RA Label T.L., McCookey Mlewiec D.J., Distche C.M., Byers P.H.;
"Thrombospondin II: partial cDNA sequence, chromosome location, and
RT expression of a second member of the thrombospondin gene family in
RT humans.";
RL Genomics 12:421-429(1992).
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L12350; AAA03703.1; -;
DR EMBL: M81339; -; NOT_ANNOTATED_CDS.
DR PIR: A42173; A42173.
DR HSP: P00740; 11XA.
DR MIM: 188061; -;
DR InterPro: IPR000561; -;
DR InterPro: IPR000884; -;
DR InterPro: IPR001007; -;
DR Pfam: PF00008; EGF_2; 1.
DR Pfam: PF00090; TSP_1; 3.
DR Pfam: PF00093; WVC; 1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS50092; TSP1; 3.
DR PROSITE: PS01208; WVEC; 1.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1172
FT DOMAIN 19 232
FT DOMAIN 318 375
FT DOMAIN 381 432
FT DOMAIN 437 493
FT DOMAIN 494 548
FT DOMAIN 549 589
FT DOMAIN 590 647
FT DOMAIN 648 692
FT DOMAIN 725 760
FT DOMAIN 761 783
FT DOMAIN 784 819
FT DOMAIN 820 842
FT DOMAIN 843 880
FT DOMAIN 881 916
FT DOMAIN 917 952
FT DOMAIN 953 1172
FT SITE 928 930
FT DISULFID 266 266
FT DISULFID 270 270
FT DISULFID 553 564
FT DISULFID 558 574
FT DISULFID 577 588
FT DISULFID 594 610
FT DISULFID 601 619
FT DISULFID 622 646
POTENTIAL.
THROMBOSPONDIN 2.
HEPARIN-BINDING (POTENTIAL).
WVEC.
TSP TYPE-1 1.
TSP TYPE-1 2.
TSP TYPE-1 3.
EGF-LIKE 1.
EGF-LIKE 2.
CALCIUM-BINDING (POTENTIAL).
EGF-LIKE 3.
TSP TYPE-3 1.
TSP TYPE-3 2.
TSP TYPE-3 3.
TSP TYPE-3 4.
TSP TYPE-3 5.
TSP TYPE-3 6.
TSP TYPE-3 7.
C-TERMINAL.
CELL ATTACHMENT SITE (POTENTIAL).
INTERCHAIN (PROBABLE).
INTERCHAIN (PROBABLE).
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.

FT DISULFID 652 665 BY SIMILARITY.
FT DISULFID 659 678 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E44C6F5 CRC64;
Query Match 2.3%; Score 143; DB 1; Length 1172;
Best Local Similarity 20.8%; Pred. No. 0.051;
Matches 105; Conservative 56; Mismatches 189; Indels 154; Gaps 29;
DB 86 CARPLRLART-----TDWTPAGSTQGVVHGPSREGFWCLNREORPGQ 128
DB 360 CASPEFVGECCPSGLHSVDGEBSMPAEMTQCSVT-----CGSGTQQNGR 406
QY 129 NCSNYTVRFLECP-----DGLRDRTERI-----NSPSPWKSCKSAACGOTGVOTRTRICL 178
DB 407 SCDVTSNLTCLGPSIQTRACSLSKCDTRIRQDGMGSHGSPWSSCSVTCG-VGNITRIRLCN 465
QY 179 AEMVSL-----GSEASEEQHCKMGQDC-----TACDLTQPMG--QYNADCD--- 217
DB 466 SPVPQMGSKNCKGSGREKACOGAPCPIDGRWSPWSPSACTVTCAGGIRETRVCSPE 525
QY 218 -----AC-----MC--ODFMLGAVSLP--GAPASGATYLLTKTRKLLTQTDSD 259
DB 526 PQYGGKACVGVQERQMKNSCPYDGLSNPCFPGACS-----SFPD 569
QY 260 GREPIPLGLCP-----DGKSLIKTKVKKFAPVLYLMPKTSLSKAATIKAEFVAERP--YVW 312
DB 570 GSWSC-GFCPVGFLGNGTCEBDLDECALVPDICEFSTSKVPCVNTQPGFHCIPCPRYRG 628
QY 313 MNPEKARRAAGSVALCKKATGKPRPDXYFHTNDLDPISLYKHEKSLVLRKLDQHQAG 372
DB 629 NQPVGVGLEAKTEKQVCEPE-NPCKDK--THN-----CHKHAECIYLGHFSDPM-- 675
QY 373 EYFCKAQSDAGAVKSKVLAQLIYTADEFPCNPVPESYLLRLPHDFONATNSFYDV--GR 431
DB 676 -YKCECQTG-----YAGDGLICG--EDSDLDGPNLNLVCATNATFYICIDN 719
QY 432 CPVKTCAGQODNGIJCRCDAVQNC-----GISKTEREIQCSGYTLPTVAKKCS--- 481
DB 720 CPHLPSNGQEDPD---KKGIDACDDDDNDNGV--TDKED-NCQLLEFNPQADYDXKEVG 773
QY 482 --CORCTETRSIVKGRVSAADNGE 503
DB 774 DRCDNCPYVHN--PAQIDTNNGE 795

Search completed: October 9, 2001, 11:53:54
Job time: 219 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2001, 11:38:59 ; Search time 26.29 Seconds
(without alignments)
2730.270 Million cell updates/sec

Title: US-09-609-383-2

Perfect score: 6317
Sequence: 1 MVTGKAWMSFLVLEVTSVL.....QSGVVASLRFPRVAQOPLIN 1184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6307	99.8	1184	21	AAV6657
2	6307	99.8	1184	22	Membrane-bound pro
3	6297	99.7	1184	20	Membrane-bound pro
4	3088.5	48.9	1156	20	AAV74445
5	2898.5	45.9	599	21	AAV23754
6	170	2.7	890	22	AAV49966
7	170	2.7	890	22	AAV49502
8	167.5	2.6	481	20	AAV50003
9	167.5	2.6	1072	22	AAV40145
10	160	2.5	905	22	AAV72300
11	157.5	2.5	1081	22	AAV72284
					Human ADAMTS-8 am
					Human ADAMTS-10 am

12	157	2.5	1152	21	AAV00042	Human thrombospond
13	157	2.5	1170	22	AAV74450	Human variant thro
14	156	2.5	680	21	AAV21251	Human metalloprote
15	155	2.5	1496	20	AAV81030	Melanoma associate
16	155	2.5	1496	21	AAV70469	Human p53 target m
17	154.5	2.4	4412	21	AAV53666	Sequence g1/101742
18	151.5	2.4	564	21	AAV94406	Human ACAM4/IgG4-F
19	151.5	2.4	598	21	AAV94404	Human ACAM4/IgG4-F
20	151.5	2.4	4393	22	AAV31889	Amino acid sequenc
21	148.5	2.4	300	21	AAV00041	Human COMP/TSP-2 c
22	147.5	2.3	1125	19	AAV52288	Rattus norvegicus
23	147.5	2.3	1139	19	AAV57779	Rattus norvegicus
24	144	2.3	1242	19	AAV52287	Rattus norvegicus
25	144	2.3	1911	16	AAV17726	Human ptp-OB. Hom
26	144	2.3	1911	18	AAV72225	Human protein tyro
27	144	2.3	1911	20	AAV94027	Human protein tyro
28	143	2.3	1172	21	AAV00043	Human thrombospond
29	143	2.3	1291	16	AAV75201	Tyrosine phosphata
30	142.5	2.3	4544	15	AAV47861	Alpha 2-Macroglobu
31	142.5	2.3	4544	15	AAV60517	Human alpha-2-MR.
32	142	2.2	1172	21	AAV19677	Human thrombospond
33	140.5	2.2	368	21	AAV94405	Human ACAM4/IgG1-F
34	140.5	2.2	398	19	AAV80405	A secreted protein
35	140.5	2.2	398	20	AAV29592	Human MBGP1 protei
36	140.5	2.2	398	20	AAV13358	Amino acid sequenc
37	140.5	2.2	398	21	AAV94403	Human ACAM cellula
38	140.5	2.2	398	21	AAV69288	Human amino acid sequenc
39	140.5	2.2	398	21	AAV45095	Human LDCAM bindin
40	140.5	2.2	398	22	AAV80226	Human PRO258 prote
41	140.5	2.2	432	21	AAV94402	Human angiogenesis
42	140.5	2.2	432	21	AAV94402	Human ACAM cellula
43	140.5	2.2	432	21	AAV9286	Amino acid sequenc
44	140.5	2.2	432	21	AAV45094	Human LDCAM bindin
45	140.5	2.2	433	21	AAV53272	Human Beat-1-like 1

ALIGNMENTS

RESULT 1	
AAV6657	standard; protein: 1184 AA.
ID	AAV6657
XX	XX
AC	AAV6657;
XX	XX
DT	05-APR-2000 (first entry)
XX	XX
DE	Membrane-bound protein PRO1188.
XX	XX
KW	Membrane-bound polypeptide: PRO polypeptide; LDL receptor; TIE ligand;
KW	pharmaceutical; receptor immunoadhesin; gene mapping.
XX	XX
OS	Homo sapiens.
XX	XX
PN	W09963088-A2.
XX	XX
PD	09-DEC-1999.
XX	XX
PF	02-JUN-1999; 99WO-US12252.
XX	XX
PR	02-JUN-1998; 98US-0087607.
PR	02-JUN-1998; 98US-0087609.
PR	02-JUN-1998; 98US-0087759.
PR	03-JUN-1998; 98US-0087827.
PR	04-JUN-1998; 98US-0088021.
PR	04-JUN-1998; 98US-0088025.
PR	04-JUN-1998; 98US-0088028.
PR	04-JUN-1998; 98US-0088029.
PR	04-JUN-1998; 98US-0088030.
PR	04-JUN-1998; 98US-0088033.
PR	04-JUN-1998; 98US-0088326.
PR	05-JUN-1998; 98US-0088167.
PR	05-JUN-1998; 98US-0088202.

are useful as hybridization probes, in chromosome and gene mapping and in the generation of antisense RNA and DNA. PRO nucleic acid sequences will also be useful for the preparation of PRO polypeptides, especially by recombinant techniques.

Sequence 1184 AA;

```

Query Match          99.8%; Score 6307; DB 21; Length 1184;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1182; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 MWGTAWWFSLVLEVTSLGKQMTQSVRRVQGGKKNPSFFAKPADLPESGEMTTLF 60
DB 1 mvgtlawwfsflvlevtslgkqmtqsvrrvpggkknpsffaapadllespgemtlf 60

OY 61 NIDYPGKGDYERLDAIRFYYGDRVCARPLRLAARTDWTMTAGSTGQVHSGPRGFNCL 120
DB 61 nidypgkgdyerldaifryygdrcarplrlaartdwtmtagstgqvvhsgprgfncl 120

OY 121 NREQRGQNCNNTVTRFLCPGSLRRDTERIWSWSPWSKCSACGQTEGVTTRICLAE 180
DB 121 nreqrgqncsnntvtrflcpgslrrdterlwspswskscsaacgqtgvttriclalae 180

OY 181 MWSLCSASESGOHCMGDDCTACDLTCPMGOVNADCMQDDEMLHGAVSLRPGAPASG 240
DB 181 mwslcsaseeegohcmgddctacdltcpmgvnaadcdacmqdcmhgvavslrpgapasp 240

OY 241 AAIYLLTTPKLLTQTDSDGRFRIRGLCPDGKSLIKITKVKAPAPVLMMPKSLAAATIK 300
DB 241 aaiylltltkplltqtdsdgrfrirgplcpdgkslikitkvkapapvlmmpkkslaaatik 300

OY 301 AEFVRAETPYWYNDETAKARRAGQSVSLCCAKYGRPRDQYFWYHNDLLDPSLYKHESK 360
DB 301 aefvraetpywymnpektarragqsvslccakgkprdqyfwyhnndlldpslykhesk 360

OY 361 LVLRKLOQHOGGEYFCRCKOSODGAVKSKVAOLITWASDETPCNVPESTLRILPHDCQN 420
DB 361 lvlrklqhqdaeyfcckqsdagavksvaqlitwlasdepcnpvpestylrphdcqn 420

OY 421 ATNSFYVVGRCPTVTCAGQODNGIRCRDAVONCCGISKTEREIQSGSYTLPTKYAKEC 480
DB 421 atnsfyvvgrcptvtcagqodngircrdavnccgiskteereiqscsytlptkyakec 480

OY 481 SCQRTETRSIYRGNVSAADNGEPMRFGHVWGNRSVSMYTGKGGFTLHVPODTERLYLT 540
DB 481 scqrtertrsiyrgnvsaadngemrfghvwmgnrsvsmtygkgtflhvpofterlylt 540

OY 541 FVDRLQKQVNTTKVLPFKKKGSAVFHEITKMLRKREPTITLEAETNITPLGEVVGEDPMAE 600
DB 541 fvdrlqkqvnttkvlpfkkkgsavfheitkmlrkreptitleaetnitplgevvgedpmae 600

OY 601 LEIPRSFYRONGEPIYKGVKASVTFLDPRNISTATAQTDFNINDEGDTPLRTYGMF 660
DB 601 leiprsfyrgngepiykgvkvastvfldprnistataqtdfnindegdtplrtymf 660

OY 661 SVDFDEVTSSEPLNAGKVKVHLDSQVAKPEHISTVKLWLSNLPDGLWEEBGEDEFENOR 720
DB 661 svdfdevtseplnagkvkvhlqskvpehistsvklwlsnlpdglweebgedfknor 720

OY 721 RKKRRDRFTLVGNLEIRERLFLNDVPESRCFVYVRAVRSRRFLPSEDIQGVISVTL 780
DB 721 rkkrrdrftlvgnleirerlflndvpesrcfvvrayvrsrrflpseidqgvvisvltl 780

OY 781 EBRPFGLSNPRAMGRFDSVITGPNACVAFCDQSPDVSAYVYVLAASLAGEELQAVESSP 840
DB 781 ebrpfglsnpramgrfdsvitgpnacvafcdqspdvayvyvylaslageelqavessp 840

OY 841 KKNPNAIGVPOPYLKNLNRFTDHDPRVYKTAFOISMAKPPNSAESNGPIYAFENIR 900
DB 841 kknpnaiigvppylknlnrftdhdprvyktafaismakppnsaesngpiyafenir 900

OY 901 ACEEAPPSAAHRRFYQIEDRDYDNTVTPNEDDPMKSWMTEDYLAAMPKPMERFACITKYKI 960

```

```

DB 901 aceeappsaahrrfyqiedrdydyntvtpneddpmkswmedylawmpkmeffracylkvi 960
OY 961 VGPLEVNVNRSRMGCTHRRYTGKLYGIRDVNRSTRDQPNVSAACLEFKCSMLYDDORV 1020
DB 961 vgplevnvnsrmgcthrrytgklygirdvnrstrdqpnvsaaclefkcsqmlyddorv 1020

OY 1021 DRTLVKVIPOGSCRRASVPMLEHYLVNRLPLAVNNDSSEYMLAPLDPLGHNGCIYWT 1080
DB 1021 drtlvkvipogscrrasvpmlehylvnrlplavnnosseymlapldplghngciywt 1080

OY 1081 DDDPTAKEIALGRCFDGTSDGSSRIMKSNVGVALTFCNVEQVQROSAFOYLOSTPAOS 1140
DB 1081 dddptakeialgrcfdgtdsgssrimksnvgvaltfcncevqyrqsafgylqgstpags 1140

OY 1141 PAAGTVQGRVPSRRQORASRGQROSGVVASLRFPRAAOPLIN 1184
DB 1141 paagtvqgrvpsrrqorasrgqrosgvvaslrfrvaqplln 1184

RESULT 2
AAB65180
ID AAB65180 standard; Protein; 1184 AA.
XX
AC AAB65180;
XX
DT 02-APR-2001 (first entry)
XX
DE Human PRO1188 (UNQ602) protein sequence SEQ ID NO:124.
XX
KM Human; secreted and transmembrane protein; PRO; cytosstatic;
KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
OS diagnostic assay.
XX
OS Homo sapiens.
XX
PN WO200073454-A1.
PD 07-DEC-2000.
XX
PE 30-MAR-2000; 2000WO-US08439.
XX
PR 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 17-AUG-1999; 99US-0149396.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 08-OCT-1999; 99US-0158663.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 15-MAR-2000; 2000WO-US05884.
PR 20-MAR-2000; 2000WO-US07377.
XX
FA (GETH ) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerlitsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WT;

```


PI Yocum S;
 XX WPI: 1999-189634/16.
 DR N-PSDB: AAX18449.
 XX
 PT New human nucleotide pyrophosphohydrolase - useful for providing
 PT methods for identifying and treating arthropies, immunological
 PT disorders, and cancer
 XX
 PS Claim 1; Fig 1; 42pp; English.
 XX
 CC This sequence represents the human nucleotide pyrophosphohydrolase,
 CC designated NPPH-1, of the invention. NPPH-1 antagonists, antibodies,
 CC agonists, proteins, complementary sequences or vectors can be used to
 CC treat and identify arthropies (e.g. calcium pyrophosphate dihydrate
 CC deposition disease, degenerative joint disease, fibromyalgia,
 CC haemochromatosis, osteoarthritis, progressive systemic sclerosis,
 CC pseudogout, psoriasis, rheumatoid arthritis and lupus erythematosus);
 CC immunological disorders (e.g. AIDS, allergies, anaemia, asthma,
 CC ulcerative colitis, dermatomyositis, diabetes mellitus, emphysema,
 CC glomerulonephritis, gout, multiple sclerosis, osteoporosis and
 CC pancreatitis), trauma; complications of cancer, haemodialysis, and
 CC extracorporeal circulation; viral, bacterial, fungal, parasitic,
 CC protozoal, and helminthic infections; and cancer (e.g. adenocarcinoma,
 CC lymphoma, melanoma, myeloma, sarcoma, leukaemia, or teratocarcinoma of
 CC the bone and bone marrow, brain, breast, cervix, gastrointestinal tract,
 CC kidney, liver, lung, ovary, testis and skin).
 XX
 SQ Sequence 1184 AA;

Query Match 99.7%: Score 6297: DB 20: Length 1184;
 Best Local Similarity 99.7%: Pred. No. 0;
 Matches 1181; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MWGTAWFESFLYEVTSTLGRQMTQSVRRVOPGKKNPSIFARPADTLESPGEMTLF 60
 DB 1 mvgtkawfslflevstslgrtmqltsvrrvpgkknpsifarpadtlespgewtlf 60
 QY 61 NIDYGGKGDYERLDALIRFYGDYDVCARPLLEARTTDMTPAGSGQVYHSGSPREGFWCL 120
 DB 61 nidyggkgydyerldalirfygydyvcarplllearttdmtpagsgqvvyhsgspregfwcl 120
 QY 121 NNEORPGONCSYTVTRPGLCPGSLRDRTERITWSPSPWKSACAGOGVOTRTRICLAE 180
 DB 121 nneqrpqgncsytvrpqlcpgsllrderltwspwskscacagqvtrtrtriclae 180
 QY 181 MWSLCSSEASEGQHMGGDCTACDLTCPCMGQVNAADCAOMCFMILHGAVSLPGGAPASG 240
 DB 181 mwslesseaseegqhmggdctacdltcpmgqvnaadcaomcfmihgavslpggapasg 240
 QY 241 AAIYLLTTPKLLTQTDSDGRFRIFGLCPDGKSLIKTRKVFAPVILMPKTSLSAATLK 300
 DB 241 aaiylltkpklltqtdsggrfrifglcpdgksliktrkvfapviltmpktslkaatlk 300
 QY 301 AEFVAERPYMMNENETARBARAGOSVSLCCKATGKPRDQKXYWYHNDLLDSLYKHNEK 360
 DB 301 aefvraeprymnenetarbaragosvslcckatgkprdkxywyhndlldslykhnek 360
 QY 361 IVLRLKLOHQAGEYFCKAQSDAGAVKSKVAQLIVTASDETPCNPPESYLIRLPHDCFON 420
 DB 361 lvrlklqhgaeyfckagsdagavkskvaqlivlasdetpcnppesylirlphdcfqn 420
 QY 421 ATNSFYDYVGRCPVKTCAGQDNGIRCRDAVONCCGISTEBERLQCSYTLPTVYAKEC 480
 DB 421 atnsfydyvgrcpvktcagqngircrdavnccgistslterlqcsytlptvyaekc 480
 QY 481 SCORCFETPSIYGRVSAADNGEPMRFGHVYMGNSRVSTGKGFETLHVPODTEBLVLT 540
 DB 481 scorcfeetpsiygrvsaadngepmrfghvygmnsrvstgkgyfletlvhvpdteblvlt 540
 QY 541 FVDRLQKFPVNTKVLFPFNKGSVAFHEIKMLRKEPITLAEAMETNIIPLGEVVGEDPMAE 600
 DB 541 fvdrlqkfpvntkvlfpfnkgsavfheikmlrkepitlaeamentniiplgevvgedpmae 600

DB 541 fvdrlqkfvntkvlfpfnkgsavfheikmlrkepitlaeamentniiplgevvgedpmae 600
 QY 601 LEIPSRSEYRQNGEPIYIGKVASVTFLEDPNRISTATAQTDLNFINDEGDFRPLTYGMF 660
 DB 601 leiprsseyrqngepiyigkvasvtfledpnristataqtdlnfinddegdfrpltygmf 660
 QY 661 SVDFDEYTSSEPLNKGKVVHLDSTQVKMPEHISVYKLSLNPDIQWMEEGEDFKFENOR 720
 DB 661 svdfdeytsseplnkgkvvhldstqvkmpeshsvyklslnpdiqwmeegeedfkfengr 720
 QY 721 RNKREDRTFLVGNLEIRERRFLNLDVPSRRCFVVRARSRERFLPSEIQGVSVIYL 780
 DB 721 rnkredrtflvgnleirerflnldvpsrrcfvvrarsrerflpseiqgvsviyl 780
 QY 781 EPRTFPLSNPRAMGRFDSYITGPNACVAPACDDSPDAISAVYLAAGELAVESSP 840
 DB 781 eprtlflsnpramgrfdsyitgpnacvafcdspdayasvylaaeelavessp 840
 QY 841 KNPNAICVPOPYLKLNVNRRTHEDPRYKTAFOISMAKPPNBSAEPNSGPIYAFENLR 900
 DB 841 knpnaicvpopylklnvnrthedpryktafoismakppnsaesngpiyafenlr 900
 QY 901 ACEEAPPSAHRFRFYQIEGDRYNTVPFENEDDPKSWTEDIYLAWMPKPMEEFRACYIKVKI 960
 DB 901 aceeappsahrfrfyqiegdryntvpfenneddpkswtediyawmpkpmefracykiki 960
 QY 961 VGPLEVNVSRNMGSTHRRYVGLKLGINDVNSTRRDQPNVSACLERCSGMLYDQDRV 1020
 DB 961 vgplevnrsmgsthrryvgllkylgindvnstrrdqpvnvsaclercsgmlydqdrv 1020
 QY 1021 DRTLVKVIPOGSCRASVNPMLHEYLNVNHLPLAVNNDSSEYFMLPLPLIGNVGIYVT 1080
 DB 1021 drtlrvkviipogscrasvnpmlheylvnvnhlplavnnssayfmllplplignvgyivvt 1080
 QY 1081 DDDPRFAKIALGRCFDGTSDGSSRIMKSNVVALTFMNCVEROVSQSAFOYLOSTPAQS 1140
 DB 1081 dddprfakialgrcfdgtdsgssrlmksnvvaltfmncverogvsafgyloqstpaqs 1140
 QY 1141 PAAGTVQGRVPSRRROORASRGOROSGVYASLRFPRVAQOPLIN 1184
 DB 1141 paagtvqgrvpsrrorasrgorosgyvaslrfprvaqopliln 1184

RESULT 4
 AAY23754
 ID AAY23754 standard; Protein; 1156 AA.
 XX
 AC AAY23754;
 XX
 DT 09-SEP-1999 (first entry)
 XX
 DE Human nucleotide pyrophosphohydrolase-2.
 XX
 KW Human; nucleotide pyrophosphohydrolase-2; NPPH-2; arthropathy;
 KW Bechet's syndrome; degenerative joint disease; lupus erythematosus;
 KW rheumatoid arthritis; psoriasis; immunological disease;
 KW acquired immune deficiency syndrome; Addison's disease;
 KW ulcerative colitis; osteoporosis; microbial infection; cancer.
 XX
 OS Homo sapiens.
 OS
 PN WO9932610-A1.
 PN
 XX 01-JUL-1999.
 PD
 XX 02-DEC-1998; 98WO-US25558.
 PF
 XX 22-DEC-1997; 97US-0996083.
 PR
 XX (INCY-) INCYTE PHARM INC.
 PA
 XX Hutchinson N, Lawton M, Magna H, Mitchell PG, Murry LE;
 PI Schaffer P, Yocum SA;

```

XX  WPI; 1999-418919/35.
DR  N-PSDB; AAX81946.
XX
XX  Human nucleotide pyrophosphohydrolase-2
PS
XX  Claim 1; Fig 1A-K; 89pp; English.
XX
CC  The present sequence represents human nucleotide pyrophosphohydrolase-2
CC  (NTPPH-2). Antagonists of the protein are used to treat arthropathy
CC  (e.g. Behcet's syndrome, degenerative joint disease, lupus
CC  erythematosus, rheumatoid arthritis, psoriasis and many others);
CC  immunological disease (e.g. acquired immune deficiency syndrome,
CC  Addison's disease, ulcerative colitis, osteoporosis, microbial
CC  infections and many others), also a wide range of cancers.
XX
SQ  Sequence 1156 AA:

Query Match 48.9%; Score 3088.5; DB 20; Length 1156;
Best Local Similarity 50.3%; Pred. No. 6.5e-271;
Matches 584; Conservative 177; Mismatches 360; Indels 39; Gaps 12;

QY 24 TMLTQSVRRVQPKKNPSIFAKPADLTLESPEWTLTFLNDIPGKGDEYERLDAIRFYGD 83
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 32 talglerrsvlytgpspal-----edweeasewtswtvhdhpgdgdfesaairfygyp 86

QY 84 -RVCARPLRLAEARTTDMTPAGSTGOVYHSPREGFWCLNREORPGONCSNTYFELCPG 142
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 87 arvcprlaleartldwalpsavgerhlnptfwtclnrepgtrcsnyhvftrcp-- 144

QY 143 SLRRDTERIMSPWSPWKSACAGQGTGVTTRICLAEMWSLCEASEEGHMGODCTA 202
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 145 -----leaswgawgwpwgcscsq-pgrlrrrhpcspagdacgrpleaqkcvrtrcpq 198

QY 203 CDLTCPMGQWADDDACMQDFMLHGAVSLPGAPASGAALYLLTKPKLLTQDSGRF 262
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 199 csl-----dtcecpdhllgsvlvpsgqpllgarvsltdq-pqtavatsdangltf 246

QY 263 RIFGLCPDGKSLIKTKYKF-APIVLIMPKSLKAATIKAEFVAENPYVMNRETARR 321
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 247 rvpvcvcdstranxaqmdgfsageaqaqngslsvtllld- klekpylvkhpsevrre 304

QY 322 AGQSVSLCCNATGKRPDKYFYWNHDLDPISLYKHESKVLRLKQHQAGEYFCKAQSD 381
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 305 agqvtfccaksgyqmpbkyswfhngtlldrahgyahlelrgldrpdqgqlyncakwne 364

QY 382 AGAYKSKVAQLIYVASETPCNPEPESYLRLPHDCFQNAFNSFYVYGRCPVTCAGQO 441
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 365 agavrsqtartvlapyqpadcprryellkllpedcsgpqsypayldvqlcptrcpsla 424

QY 442 DNGIRCDAYONCCGISKTEEREIQCSGYTLPTVAKESCQRCETRSIVRGVSAADN 501
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 425 gssprcgdaasrcscvrrlerelhcpgyvlrvkvaecgqkclprglvrgvvaads 484

QY 502 GEPKRFCHVYMGNSRVMTGKGTFLHVPQDTERLVLYFEDRLQCKVNTTKVLPENKKG 561
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 485 gepirfarillgqepidftaygdgftievpsqtqlvltvfvdpsgfmdavrvlfpdgr 544

QY 562 SAVHEIKMLRRKEPITLLEMETNIPLGEVYGEDPAELEIPRSRYRQGEERYICKVK 621
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 545 agvyhevkamkkapvllhsqsnclprlgeledeaplgelylpsqatrradqkpsypve 604

QY 622 ASVYFLDPRLNSTATAQTDLNFINDGDFPLRTYGMFSVDFDEYTSBPLNKGKVVH 681
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 605 arvfvdprdltsaaspsdlrfdvsdgelaprltygmfsvdlrapsaaqlygypavtr 664

QY 682 LDSTQVMPHEHISTVKIWSLNPDTGLWEEDGFEENOR--RNKREDRTFLVGNLEIRER 739
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 665 vaasqllmpgihvaekllwslmpetglweesggrfregsgprvreevflvgnveiter 724

QY 740 RLFLMDVPESSRCFVKVRAVYSEKFLPSEQGVIVINLEPRTGLSNRAAGRDSV 799
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

DB 725 rlfndvperirrcfvkrayandkftpseqevyvvllvnllepapgfspanprawgrfda 784
QY 800 ITGPNGACVPAFCDDQSPDASAYVLASLAGEELQAVESSKFFPNKAGVQPYLKNLY 859
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 785 vtgpnagclpafcdadpdaytalvtallggeelepapslprlpayvtyqpldrlgy 844

QY 860 RRTDHEPVRVKTAFOISMARPNBSAESNGPIYAFENLRACEAPPSAAHFRFOIEG 919
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 845 rrtdddpakfrngfrlnlakprgdpdaeanqpyvwrslrecgapytashffarvea 904

QY 920 DRXYNTVYFHEDDPMSTDEYLAWPKPMFRACIKVKIVGLEVNVSRRMNGHRR 979
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 905 dkyevnvprfregtpaswtqllawmpnpqeltracflkvklygqeywvtrshnagshpr 964

QY 980 TVGKLYGIRDVSRDRDQPVNSAACLEFKSGMLYODRDVRLVYVIFQSGCRSAVN 1039
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 965 trgqllygldarsvrdpserpitsaacrefkcsnmlfdqrvdrllvclmpgscrrtav 1024

QY 1040 PMLHEVLYVNLPLAVNNDTSEYTMALAPLDLGHNYGIYTVTDQPRAKETALGRCPGT 1099
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1025 glldyltrhppvpaeedpaafsmldpdyghnygytyvldgsprlaketaigcfdgs 1084

QY 1100 SDGSSRTMKSNVGVALTENCYERQVQSAFOYLQSTPQASPAAGTVQGRVPSRQORAS 1159
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1085 sdgfsremkdaqtaavlftqcreppgrpslfrllsespa--talgydltr-----remseaa 1137

QY 1160 RGGQROSGVVASLRFPRVAQ 1179
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1138 qagarsagpilttr-grvtrg 1156

```

```

RESULT 5
AAY49966
ID AAY49966 standard; Protein; 599 AA.
XX
XX AAY49966;
AC
XX
XX 03-FEB-2000 (first entry)
DT
XX
DE Porcine chondrocyte nucleotide pyrophosphohydrolase.
XX
XX Porcine; pig; chondrocyte; nucleotide pyrophosphohydrolase; NTPPH;
KW hydrolysis; nucleoside triphosphate; ATP; nucleoside monophosphate;
KW ppi; calcium pyrophosphate dihydrate; CPPD; articular cartilage;
KW age; articular tissue degeneration; arthritis; joint tissue breakdown;
KW vesicle formation; diagnosis; cartilage.
XX
XX Sus scrofa.
OS
XX U05986080-A.
PN
XX 16-NOV-1999.
PD
XX 17-OCT-1997; 97US-0954333.
PE
XX 18-OCT-1996; 96US-0028702.
PR
XX
XX (MCWR-) MCW RES FOUND INC.
PA
XX
PI Masuda I, Halligan BD, McCarty DJ, Ryan LM, Barbieri JT, Haas AL;
XX WPI; 2000-022330/02.
DR
XX N-PSDB; AAZ35852.
PS
XX Porcine nucleotide pyrophosphohydrolase coding sequences and protein
XX
XX Disclosure; Column 19-24; 25pp; English.
XX
XX The present sequence represents nucleotide pyrophosphohydrolase (NTPPH)
XX isolated from a Porcine chondrocyte cDNA library. NTPPH catalyzes the
XX hydrolysis of nucleoside triphosphates, such as ATP, to a nucleoside
XX monophosphate and Ppi. This is an important factor in the generation
XX of Ppi, which then forms calcium pyrophosphate dihydrate (CPPD). CPPD

```


CC crystal deposition in human articular cartilage is a common age-dependent
 CC event often associated with degeneration of articular tissues and/or with
 CC acute attacks of arthritis. Knowledge of the enzymes involved in this
 CC process is important for combating, e.g. arthritis. Sufficient amounts
 CC of the enzyme would also be useful for allowing crystallisation to
 CC determine the enzyme structure. It can also be used as an antigen to
 CC raise antibodies useful in diagnosis as a marker of joint tissue
 CC breakdown and of vesicle formation within articular tissues. The cDNA
 CC can also be used in in situ hybridisation to determine the presence and
 CC location of mRNA for the parent molecule within cartilage and other
 CC tissues. The cDNA of NPPH can be used to generate the protein for
 CC use in screening methods for modulators of the enzyme.

Sequence 599 AA;

Query Match 45.9%; Score 2898.5; DB 21; Length 599;
 Best Local Similarity 91.1%; Pred. No. 3.8e-254;
 Matches 544; Conservative 23; Mismatches 29; Indels 1; Gaps 1;

OY 588 PLGEVVGEDPMALFLPSPSRKONGEPKVKASVTFLLDRNRINSTATAOITLNFIND 647
 DB 4 PLGDMGDEGPMGELPSPKSYRNGEPYTGKVASVTFLLDRNISTATAQSDINFIN 63
 OY 648 ECDTFPLRTYGMFVDFRDEVTSEPLNAGKVHLDSIOVKMPEHISTVKLSLNDPTGL 707
 DB 64 EGDTPITLYGTFVDFIDEAASESLNVGKVVHDSICQKMEIHPMMKLVSLNPDGL 123
 OY 708 WEEEDGFKEFNGRRKREDRTFLVGNLEIRERRLFNLDVPSRRCFVKVRAYSERFLPS 767
 DB 124 WEEEDGFKEFNGRRKREDRTFLVGNLEIRERRLFNLDVPSRRCFVKVRAYSERFLPS 183
 OY 768 EIOGVVSVITLLEPRTGLSNPRAMGRDSYITGNACVAFDDSDPAKSAVYLAS 827
 DB 184 EGIQGVVSVITLLEPRTGLSNPRAMGRDSYITGNACVAFDDSDPAKSAVYLAS 243
 OY 828 LAGEELQAVESPPKFNPAIGVOPYLKLNLRTRDHEPDKTAFOISMAKPPNSAE 887
 DB 244 LAGEELQAVESPPKFNPAIGVOPYLKLNLRTRDHEPDKTAFOISMAKPPNSAE 303
 OY 888 ESNGRPIYAEFLRACEEAPSAHFRFYQIEGDRYDNTVPNEEDPMSTEDYLAWMPK 947
 DB 304 ESNGRPIYAEFLRACEEAPSAHFRFYQIEGDRYDNTVPNEEDPMSTEDYLAWMPK 363
 OY 948 PHEFACVYKVIKIVPLENVSRMNGHRRRTVKLYGIRVRSRRDDQNNVSAQLE 1007
 DB 364 PHEFACVYKVIKIVPLENVSRMNGHRRRTVKLYGIRVRSRRDDQNNVSAQLE 423
 OY 1008 FKCSGMLDODRVDRLVIVPQSCRRASVNPMLHEYLVNHLPLAVNNDTSEYTLAPL 1067
 DB 424 FKCSGMLDODRVDRLVIVPQSCRRASVNPMLHEYLVNHLPLAVNNDTSEYTLAPL 483
 OY 1068 DFLGHNNGIYVTDDPRTAKEIALGRCEFDGSDSSIRMSKSNVGVALLPNCVEROVGRQ 1127
 DB 484 DFLGHNNGIYVTDDPRTAKEIALGRCEFDGSDSSIRMSKSNVGVALLPNCVEROVGRQ 543
 OY 1128 SRFQYLGSRPASPAGVYQGVPSRQORASRGORGSGVVASLRFVPAQOPLIN 1184
 DB 544 SRFQYLGSRPASPAGVYQGVPSRQORASRGORGSGVVASLRFVPAQOPLIN 599

RESULT 6

AA49502 standard; Protein: 890 AA.

AA49502;

10-JAN-2000 (first entry)

Human METH2 protein.

Human: METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;

cancer; diagnosis; hyperproliferative disorder; autoimmune disease;

KW angiogenesis inhibitor; abnormal wound healing; inflammation;
 KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
 KW diabetic retinopathy; macula degeneration; haemangioma; detection;
 KW arterial-venous malformation; immune deficiency.

OS Homo sapiens.

PN W09937660-A1.

PD 29-JUL-1999.

PF 22-JAN-1999; 99WO-US01313.

PR 23-JAN-1998; 98US-0072298.

PR 28-AUG-1998; 98US-0098539.

PA (IRUE/) IRUELA-ARISPE L.

PA (HAST/) HASTINGS G A.

PA (RUBE/) RUBEN S M.

PI Irue-la-Arispe L, Hastings GA, Ruben SM.

DR WPI; 1999-590684/50.

DR N-PSDB; AA232001.

PT New isolated metalloprotease thrombospondin polypeptides, useful for

PT treating hyperproliferative disorders, cancers or autoimmune disorders

PT Claim 10; Fig 2; 457bp; English.

CC AA232000 and AA232001 encode, and AA49501 and AA49502 represent, human

CC metalloprotease thrombospondin (METH) proteins METH1 and METH2

CC respectively. METH1 and METH2 have been found to be potent inhibitors of

CC angiogenesis both in vitro and in vivo. They can be used for treating

CC cancer and other disorders related to angiogenesis including abnormal

CC wound healing, inflammation, rheumatoid arthritis, psoriasis,

CC endometrial bleeding disorders, diabetic retinopathy, some forms of

CC macula degeneration, haemangiomas, and arterial-venous malformations.

CC They may be useful in treating deficiencies or disorders of the immune

CC system, by activating or inhibiting the proliferation, differentiation,

CC or mobilization (chemotaxis) of immune cells. The etiology of these

CC immune deficiencies or disorders may be genetic, somatic, such as

CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or

CC toxins), or infectious. They can also be used to treat inflammatory or

CC conditions, both chronic and acute conditions. The products can also be

CC used for detection and diagnosis. AA232002 to AA232080, and AA49503 to

CC AA49511 represent sequences given in the exemplification of the present

CC invention.

CC Sequence 890 AA;

Query Match 2.7%; Score 170; DB 20; Length 890;

Best Local Similarity 21.4%; Pred. No. 7.3e-06;

Matches 110; Conservative 49; Mismatches 224; Indels 132; Gaps 25;

OY 50 LESPGEWTLTFNIDYPGGKDYERLDAIRFYGD--RVCARPLREARTTWTGASTGQ 107

DB 427 Ldapgaalpl-ptlgprymalyqldqgrqrlfgrpfhrhpnatsaqdvcqlwchtdgaep 485

OY 108 VHGSPREFGWLNDQRGQNCNMYTRFLCPSSLKRDTRT-----WSPSPMSK 160

DB 486 lchtkngslpwadgicpypghlcsesg---clp---eeveerphrvydgwagpawpwe 538

OY 161 CSAAGGOTVOTRTRICLAEMWVSLCSEASESGOHCWG-----ODCTACDLTCPCMGQVAND 215

DB 539 cartcg-gvgvqfshckcd-----pepunggrlyclgrrakyschtee---cpqd--gks 587

OY 216 CDACMCODFMLHGAVALSGAPASGAATYLLTKRPLKLTQTDSDGRFRIPGLC-PDGKSI 274

DB 588 fregqcekyuaynydmdgn-----llqvwpxkygvgvprd---icklfcrtargrse 635

```

QY 275 LKIRKVKAPAPVLVIMPKPSTLKAATIKAEFVAPVYVWMPDE--TKARRAGQSVSLCCA 3322
Db 636 fkvteakvldgtllgspet--laleivrgcvkagcdhvdsprklcdkgvsgygymscrkv 6533
QY 333 TGKPRPDKIFYWYHND-----TLLD-----PSLYKHESKLVLRKLGQHQAERYFCCA 3789
Db 694 sgsltpthy--gyndivrlipagatnldvkgtrshpygndgnylalktad---gqylng 7477
QY 379 QSDAGAVKS-----KVQLLVRTASDEPCMPVESTLIRL---PHHCF----- 4188
Db 748 nlalsaleqdllyvgtklksygsatlerlqsfripdeplvtqlllyvpgefpdkvkytf 8077
QY 419 -----QNAITNSFY-----DVGHCKPVKTCAGAGDNGIRCD-----A 4500
Db 808 fvpndvfsmgskerratnllqpllnagwlgdmsecssctcgagwgrttvecrdpsgga 8677
QY 451 VQNCGISKTEERETIQCSGYLLPRKVKAKESQNC 485
Db 868 satcnkalpkpd-----akpesqic 888

```

RESULT	7
ID	AAB50003
XX	AAB50003 standard; Protein; 890 AA.
AC	AAB50003;
XX	
DT	19-MAR-2001 (first entry)
XX	
DE	Human METH2.
XX	
XX	Human; METH2; metalloprotease; thrombospondin; angiogenesis inhibition;
KW	Cancer therapy; benign tumour; ocular angiogenic disease;
KW	Rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW	Vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW	Scleroderma; trichoma; vascular adhesion; myocardial angiogenesis;
KW	Coronary collateral; cerebral collateral; arteriovenous malformation;
KW	Ischaemic limb angiogenesis; Osler-Webber syndrome;
KW	Plaque neovascularisation; telangiectasia; haemophilic joint;
KW	Angiodiroma; fibromuscular dysplasia; wound granulation;
KW	Crohn's disease; atherosclerosis; birth control.
XX	
OS	Homo sapiens.
XX	
PN	WO200071577-A1.
XX	
PD	30-NOV-2000.
XX	
PE	25-MAY-2000; 2000WO-US14462.
XX	
PR	25-MAY-1999; 99US-0318208.
PR	20-JUL-1999; 99US-0144882.
PR	10-AUG-1999; 99US-0147823.
PR	13-AUG-1999; 99US-0373658.
PR	22-DEC-1999; 99US-0171503.
PR	22-FEB-2000; 2000US-0183792.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PA	(SMIK) SMITHKLINE BEECHAM CORP.
PA	(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA	(IRUE/) IRUELA-ARISPE L.
PA	(HAST/) HASTINGS G A.
PA	(RUBE/) RUBEN S M.
PA	(JONAK/) JONAK Z L.
PA	(TRUL/) TRULLI S H.
PA	(FORN/) FORNWALD J A.
PA	(TERR/) TERRETT J A.
XX	
PI	IrueLa-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI	Fornwald JA, Terrett JA;
XX	
XX	WPI: 2001-025136/03.
XX	
XX	N-PSDB; AAC90058.

XX		
PT	METH1 and METH2 polynucleotides and encoded polypeptides, used to	
PT	inhibit angiogenesis in the treatment of disorders such as cancer,	
PT	rheumatoid arthritis and psoriasis -	
XX		
PS		
XX	Claim 15; Fig 2; 768bp; English.	
XX		
CC	The present sequence is human METH2 (ME for metalloprotease and TH for	
CC	thrombospondin). METH2 can be used for inhibiting angiogenesis in an	
CC	individual, and for treating cancer, benign tumours, an ocular angiogenic	
CC	disease, rheumatoid arthritis, psoriasis, delayed wound healing,	
CC	endometriosis, vasculogenetic, granulatioms, hypertrophic scars, nonunion	
CC	fractures, scleroderma, trachoma, vascular adhesions, myocardial	
CC	angiogenesis, coronary collateralals, cerebral collateralals, arteriovenous	
CC	malformations, ischaemic limb angiogenesis, Osler-Weber syndrome, plaque	
CC	neovascularisation, telangiectasia, haemophilic joints, angiodibroma,	
CC	fibromuscular dysplasia, wound granulation, Crohn's disease or	
CC	atherosclerosis. METH2 can also be used in birth control. METH2 can also	
CC	be used in diagnostic methods for the prognosis of cancer.	
XX		
XX		
Sequence	890 AA:	

Sequence 890 AA;

Query Match	2.7%	Score 170;	DB 22;	Length 890;
Best Local Similarity	21.4%;	Pred. No. 7.3e 06;		
Matches 110; Conservative	49;	Mismatches 224;	Indels 132;	Gaps 25

QY	50	LESGEWTLENNIDYPCGKGDEYELDAIRFYGG--RVCARPLRLKARTTDWTPAGSTGQ	107
Db	427	ldapgaalpl-ptjlgymalyqldqgcrqlfpgdfhncntsagdcqaqlwchtgdaep	4855
QY	108	VHGSPEBGEWCLNRQRPCGNGSNYVRELCPGSLKRDTERI-----WSWFSWISK	160
Db	486	lchknsgslpwadqltccpgphlcsesg----clp-----eeverpkpvddgawapwbpwge	538
QY	161	CSAAGCGTGVOTRRIRICLAENVSLCSASEEGHCHMG-----ODCTACDILTCPGGYNAD	215
Db	539	csrtcg-gyqgfsnreckd-----pepgngryclrrtrakyschnee--cpdd--gks	587
QY	216	CDACMCDPFMLHGAVSLPGAPASGAAILYLTKTPTKLLTGTDSGRPRIPGLC-PDGKSI	274
Db	588	fregqckeyanaynltdmagn-----llwqwpkyavgsprcd---rcklfcargrise	635
QY	275	LKIRKVEARPIVLMPTSTLSKAATIKAEFPAERPPVYWMPE--TKARRAGQSVSLCCKA	332
Db	636	fkvtreakvldgtlcpget--laicvrgycvkaagcdhvwdsprklldkcgvcgqkyonscrkv	653
QY	333	TGKRPRPKYEWYHND-----TLLD--PSLYKHESKILYLRKLQOHQAGEYFCKA	378
Db	694	sgsltpciny--gynditvtipagatnldvkgfshpyqndgnylalktad---qgylng	747
QY	379	QSDAGAVKS-----KVAQLVTASDELPCNPPESTLIRL---PHNCF-----	418
Db	748	nlaisalegdllvkgtllkysgsiatlerlqsfripredplvtqglllvpgavfppkvkylf	807
QY	419	-----QATNSFEY-----DVGCSPVKTCAAGODNCRKRD-----A	450
Db	808	fvpdvdfmsgskkeratnllqpllhqawylgwscscstcgagwgrttrvecrdpdsga	867
QY	451	VONCGGSKITREERITQCSGTLTPKVAKESQQRG	485
Db	868	satcnkalpked-----akpesqdlc	888

RESULT	8
AAV04145	
ID	AAV04145 standard; Protein; 481 AA
AC	XX
DT	AAV04145;
XX	
XX	15-JUN-1999 (first entry)
DE	Rat Tango-76 protein.

XX ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
 KW tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;
 KW Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;
 XX metastasis; embryogenesis; egg implantation; ADAMTS-10.
 OS Homo sapiens.
 XX MO200111074-A2.
 XX 15-FEB-2001.
 XX 03-AUG-2000; 2000WO-US212223.
 XX 06-AUG-1999; 99US-0369364.
 XX (CLEV-) CLEVELAND CLINIC FOUND.
 XX (APTE/) APTE S S.
 XX (HORS/) HORSKAINEN T L.
 XX (HIRO/) HIROHATA S.
 XX Apte SS, Hurskainen TL, Hirohata S;
 XX WPI: 2001-159978/16.
 XX N-PSDB; AAF63445.
 XX Murine and human 'A Disintegrin-like And Metalloprotease domain with
 PT Thrombospondin type I motifs' proteins and the nucleic acids encoding
 PT them, useful for treating e.g. tumours, inflammation and arthritis -
 PS Claim 1; Fig 9; 181pp; English.
 XX This invention relates to murine and human ADAMTS-N (A disintegrin-like
 CC and metalloprotease domain with thrombospondin type I motifs) proteins,
 CC designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the
 CC invention are cDNA sequences encoding the proteins, and antibodies
 CC specific for the proteins. The nucleic acid sequences and proteins may be
 CC used in the prevention, diagnosis and treatment of diseases associated
 CC with inappropriate ADAMTS-N expression. Disorders that may be treated
 CC using the nucleic acids, proteins and antibodies include, for example
 CC tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos
 CC syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage
 CC in arthritic (both inflammatory and non-inflammatory) disease,
 CC angiogenesis, tumour growth and metastases, and they may also be used for
 CC controlling embryogenesis and implantation of fertilised eggs. The
 CC present sequence represents human ADAMTS-10.
 CC
 CC Sequence 1081 AA:
 SO
 Query Match 2.5%; Score 157.5; DB 22; Length 1081;
 Best Local Similarity 20.2%; Pred. No. 0.00014;
 Matches 98; Conservative 62; Mismatches 180; Indels 145; Gaps 25;

OY 107 QVHGSPREGFECLEINREOR-----PGQNCNVTV-----RFLCPPGSLRDPER 150
 DB 466 qckygveveselwclsknrcinsipaaegtlcqhhtldkwyckrvvcyfrigsrpegydg 525
 OY 151 IWSPMSPKSCSAACGGQGVQTRTRICLAEMVSLCSEASEGQHCQMGQ-----DCTACDL 205
 DB 526 awgptwtpwgdcstcrg-gvsvsssrhcdsprptl-----ggkylgerrrrrnsctnd- 577
 OY 206 TCFMGQVNAADCACMGQDFMLHGANVSLP-----GGAA-----S 239
 DB 578 -cpbg--sqdfrevqcaef-----dsipfrgkyfwktyrggvkacsltslaegfnfyt 629
 OY 240 GAATVILRTTP-----KLTQFTDSGDGRFRIJGLCPDCKSLIKITK 279
 DB 630 eraaavogtprpdtvdlcvsgeckhvgcdvlsdlredkcrvcg--gdsaacetleg 687
 OY 280 VKFADI-----VLTMPKTSLKAAITKAEFVRAETPYVMVNPETKARRAQOSVSLCK 331
 DB 688 v-fspaspagaygedvwmvlpkgsvh---iffqdlnslshlalk-----gdqsl1lle 735

OY 332 A-TGKPRDKYFWYHNDLFLDPSLYKHKESKLVLRKLOOHQAGEYFCKAOSDAGAVKSKYA 390
 DB 736 glpypfpq-----hrlplagttffqlrpgdpdqvslea-1gpnasli 776
 OY 391 OLIVTASDETPCNVPESYLRLPHDFOMATNSFY- DVGRCPVKTCAGQODNGIRCRD 449
 DB 777 -vmvlarlelpa---lryfnaplardslppyswhypwckcsaqgsgqvqavecrn 831
 OY 450 AVQN-----CCGISKEEREIOCSGTYLTPK--VAKECSQRCQE-----TRSIYGR 495
 DB 832 qldsavaphycsahsklpkrqacntecppdpwvgnwslscsdagvistsvvcqr 891
 OY 496 VSAAD 500
 DB 892 vsaae 896

RESULT 12
 AAB00042
 ID AAB00042 standard; Protein: 1152 AA.
 XX
 AC AAB00042;
 XX
 DT 08-NOV-2000 (first entry)
 XX
 DE Human thrombospondon-1 (TSP-1).
 XX
 KW TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
 KW thrombospondin; angiogenesis; tumour; treatment; cancer;
 KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
 KW glaucoma.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX FH Region 361..416
 XX FT Region /label= Type 1 repeat region
 XX FT Region 417..473
 XX FT Region /label= Type 1 repeat region
 XX FT Region 474..530
 XX FT Region /label= Type 1 repeat region
 XX
 PM WO200044908-A2.
 XX
 PD 03-AUG-2000.
 XX
 PF 01-FEB-2000; 2000WO-US02482.
 XX
 PR 01-FEB-1999; 99US-0118053.
 XX
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PI Lawler JW;
 XX WPI: 2000-514823/46.
 XX
 CC New nucleic acids are described which encode a protein comprising
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
 CC but not the TGF (transforming growth factor)-beta activation region
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
 CC the second and third type-1 repeats and the COMP (cartilage
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
 CC caused inhibition of the growth of tumours in mice models.
 CC Thus the nucleic acids and proteins may be useful for treating
 CC angiogenesis related diseases such as cancer (by reducing the rate of

CC growth and size of tumours), arthritis, psoriasis, diabetic
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be
 CC used for treating human immunodeficiency virus (HIV) infection.
 CC Anti-angiogenic therapy has little toxicity, does not require the
 CC therapeutic agent to enter tumour cells or cross the blood-brain
 CC barrier, controls tumour growth independently of growth of
 CC tumour cell heterogeneity, and does not induce drug resistance.
 XX
 XX Sequence 1152 AA:

Query Match 2.5%; Score 157; DB 21; Length 1152;
 Best Local Similarity 20.1%; Pred. No. 0.00017;
 Matches 111; Conservative 37; Mismatches 150; Indels 254; Gaps 27;

```

QY 10 SFVLEVTSLVGRQMTLQSVRRVOPGKKNPSIFAKPADTLESP-----GEW 56
   | : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 260 ssmvlelrglrltlvtlqdsirkyteenk-----elanelirpplcyhngvyrmeeew 313
QY 57 TT-----LFNIDYPGKGDYERLDAIRFYGDRCARPLRL 92
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 314 tvdsctechcqnsvtlckkvsclmpcsnatvpdgc-----ccprcwps 357
QY 93 EARTDWTTPAGSTGQVVHGSFREGFW-----CLNRPORPGONCSNTVRFCLPPGSLRR 146
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 358 dsadddgswpse-----wtscstscgnglqgrgrscdslnmr--cegssvql 402
QY 147 DTERI-----WSPWSPWSKCSAACGQTGQVOTRTRICLAEMVSL-----CSEASE 190
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 403 rtchlgcdkrtfkqdgswshwspsscsvctcgd-gvlttrilcnspqmgkpcgear 461
QY 191 EGOHCKMGQDC-----TACDLITC-----PMGOVNA-DC-----D 217
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 462 etkackkdaaplmgwgpwspwdicavtscggyqktrslcnmpfpqfgykdcvgdvlenq 521
QY 218 ACMQODEFMLHGAVSLPGGAPASGAATVLLKTPKRLTQTDSDGRFRIPGCPDGKSLTKI 277
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 522 icnkqdcpldgclsnpcfaqyk-----ctsy-----dgsawk-gacppgys----- 562
QY 278 TKVKFADIVLTMPKTSLSKAATIRAEFVRAETPYVMNPETKARRAQSVSLCKATGKPR 337
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 563 -----gnglqcltdvdeckev 577
QY 338 PDKTFWYHNDTLDDPSLYKHESKLVLRKLOOHQAGEYFCKAQSADG----- 383
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 578 pdacfnhn-----gehrce-nltpgyncjpcprrfsgq 610
QY 384 -----AVKSKVAQLIVTASDETPCNPVPESYLIRLPHDCFONATNSF---YYD-VGRCPVK 435
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 611 pfqggyvehatankqvc-----kprnctdg-----thcnknakcnylghysdpmryceek 661
QY 436 TCAGQODNGIIRC 447
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 662 --pgyagngjlic 671

```

RESULT 13
 AAB74450
 ID AAB74450 standard; Protein; 1170 AA.

AC AAB74450:
 DT 06-JUN-2001 (first entry)
 XX Human variant thrombospondin 1.

XX Human; variant thrombospondin 1; variant thrombospondin 4; SNP;
 KM polymorphism; vascular disease; coronary artery disease; forensics;
 KW myocardial infarction; atherosclerosis; stroke; venous thromboembolism;
 XX pulmonary embolism; paternity test.
 OS Homo sapiens.
 XX

PN W0200118250-A2.
 XX 15-MAR-2001.
 PD 07-SEP-2000; 2000MO-US24503.
 XX 10-SEP-1999; 99US-0153357.
 PR 26-JUL-2000; 2000US-0220947.
 PR 16-AUG-2000; 2000US-0225724.
 XX

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA- (MILL-) MILLENNIUM PHARM INC.

PI Lander ES, Gargill M, Ireland JS, Bolk S, Daley GQ, McCarthy JT;
 XX N-PSDB; AAF95238.
 DR WPI: 2001-226749/23.

PT Nucleic acids comprising single nucleotide polymorphisms, useful in
 PT applications such as forensics, paternity testing, medicine, genetic
 PT analysis and phenotype correlations to diseases such as diabetes and
 PT atherosclerosis -

PS Claim 20; Fig 1; 242pp; English.

XX The present invention provides a method of diagnosing a vascular disease
 CC in an individual, involving determining the sequence at various
 CC polymorphic sites within the human thrombospondin 1 and thrombospondin 4
 CC genes. The sequences at a number of polymorphic sites are also provided
 CC in the specification. In particular, the method can be used in the
 CC diagnosis of atherosclerosis, myocardial infarction, coronary heart
 CC disease, stroke, peripheral vascular diseases, venous thromboembolism
 CC and pulmonary embolism. Single nucleotide polymorphisms (SNPs) are also
 CC useful in forensics, paternity testing, genetic analysis and phenotype
 CC correlations to diseases. The present sequence is the human variant
 CC thrombospondin 1 protein.

XX Sequence 1170 AA:

Query Match 2.5%; Score 157; DB 22; Length 1170;
 Best Local Similarity 20.1%; Pred. No. 0.00018;
 Matches 111; Conservative 37; Mismatches 150; Indels 254; Gaps 27;

```

QY 10 SFVLEVTSLVGRQMTLQSVRRVOPGKKNPSIFAKPADTLESP-----GEW 56
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 278 ssmvlelrglrltlvtlqdsirkyteenk-----elanelirpplcyhngvyrmeeew 331
QY 57 TT-----LFNIDYPGKGDYERLDAIRFYGDRCARPLRL 92
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 332 tvdsctechcqnsvtlckkvsclmpcsnatvpdgc-----ccprcwps 375
QY 93 EARTDWTTPAGSTGQVVHGSFREGFW-----CLNRPORPGONCSNTVRFCLPPGSLRR 146
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 376 dsadddgswpse-----wtscstscgnglqgrgrscdslnmr--cegssvql 420
QY 147 DTERI-----WSPWSPWSKCSAACGQTGQVOTRTRICLAEMVSL-----CSEASE 190
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 421 rtchlgcdkrtfkqdgswshwspsscsvctcgd-gvlttrilcnspqmgkpcgear 479
QY 191 EGOHCKMGQDC-----TACDLITC-----PMGOVNA-DC-----D 217
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 480 etkackkdaaplmgwgpwspwdicavtscggyqktrslcnmpapqfgykdcvgdvlenq 539
QY 218 ACMQODEFMLHGAVSLPGGAPASGAATVLLKTPKRLTQTDSDGRFRIPGCPDGKSLTKI 277
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 540 icnkqdcpldgclsnpcfaqyk-----ctsy-----dgsawk-gacppgys----- 580
QY 278 TKVKFADIVLTMPKTSLSKAATIRAEFVRAETPYVMNPETKARRAQSVSLCKATGKPR 337
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 581 -----gnglqcltdvdeckev 595
QY 338 PDKTFWYHNDTLDDPSLYKHESKLVLRKLOOHQAGEYFCKAQSADG----- 383

```

```

Db 596 pdact-nhm-----gehrce-ntdpgynclpccprftsq 628
OY 384 ---AVKSKVAQLVTASDETPCNPVPESYLRLPHDFCFONATNSF---YVD-VGRCPVK 435
Db 629 pfqgvgvhenatankayc---kprncptcdg-----chdcnkakcnylghysdpmryrceek 679
OY 436 TCAGGQDNGIRC 447
Db 680 --pyyagngllc 689

RESULT 14
AAB21251
ID AAB21251 standard; Protein: 680 AA.
XX AAB21251;
XX
XX 23-FEB-2001 (first entry)
XX
XX Human metalloproteinase ADAMTS-2.
XX
XX Human; ADAMTS2; metalloproteinase; ADAM;
XX a disintegrin and metalloproteinase domain; thrombospondin domain;
XX vaccine; neurotropic; neuroprotective; antiparkinsonian;
XX cerebroprotective; cytoskeletal; antiarthritic; immunosuppressive;
XX Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
XX autoimmune disease; brain tumour; brain injury.
XX
XX Homo sapiens.
XX
XX WO200053774-A2.
XX
XX 14-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US06237.
XX
XX 08-MAR-1999; 99US-0264585.
XX
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
XX PI Kelnner GS, Clark M, Maki RA;
XX
XX WPI: 2000-594326/56.
XX
XX N-PSDB: AAA95821.
XX
XX Polynucleotide encoding novel members of a disintegrin,
XX metalloproteinase and thrombospondin domain protein family used to
XX prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX
XX Claim 12; Fig 2; 129pp; English.
XX
XX The present sequence is human metalloproteinase ADAMTS-2. The
XX ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
XX and Metalloproteinase Domain) family. Members of the ADAMTS family
XX contain a thrombospondin domain in addition to the disintegrin and
XX metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
XX useful for the manufacture of medicaments for treating conditions
XX associated with neuroinflammation and/or neurodegeneration, such as
XX Alzheimer's disease, Parkinson's disease and stroke. They are also
XX useful for treating conditions associated with cell proliferation, cell
XX migration, inflammation and/or angiogenesis, such as cancer, arthritis
XX and autoimmune diseases. They can be used to treat patients afflicted
XX with an invasive tumour, a brain tumour or brain injury.
XX
XX Sequence 680 AA:
XX

```

```

Query Match 2.5%; Score 156; DB 21; Length 680;
Best Local Similarity 21.2%; Pred. No. 8.5e-05;
Matches 97; Conservative 45; Mismatches 181; Indels 134; Gaps 23;
OY 118 WCLNREQRGQCSNVTYRFL---CPGSL-----RRTERT-----WSPWSPW 158

```

```

Db 267 wchtdgaeplichtkngslpwadgtlpcgphlcsesclpveeverpkpvydgwawpawpw 326
OY 159 SKCSACQGTGVQVTRTRICLAEWVSLCSSEASEGHCNMG-----ODCTACDILTCPMGVN 213
Db 327 gescrtcg-gyqfqbhrekd-----pepqngryclgrtakyschtee--cpdd--g 375
OY 214 ADCDACMKODEFHLHGAVSLPGGAPASGAATYLLTTPKLLTQTDSDGRFRIPGLC-PPGK 272
Db 376 ksfreqqcekyanlyltmdgn-----llqvwpxyagvsprd---rcklforargr 423
OY 273 SILKITTKVFAPIVLMDPTSLKAATIKAEFVRAETPYWVMN--PETKARRAGQSVSLCC 330
Db 424 sefkvfeakvldgtlcspec--lalcvtgqcvkagcdhvdsfwkldkcygcgkqnsctr 481
OY 331 KATGKRPDPKYFWYHND-----TLTD-----PSLYKHESKLVLRKLOHQAEGEFC 376
Db 482 kgsqsltpcny--gynndlvtlpagatnldvkqshpnyqndgnylalktad----gyll 535
OY 377 KASDAGAVKS-----KVAQLIVTASDETPCNPVPESYLRL----- 413
Db 536 ngnlalsaleqdlvkgtllkysgsiatlerlqsfrrlpeplvtvqllavpgevfpkvky 595
OY 414 ----PHD-----CFONATNSFY-----DVGRCVPKTCAGGQDNGIRCRD--- 449
Db 596 ttfvprndvdfsmgskeatcniitqpllhaqvwlgdwsecscgcgawqrrtvecdpsg 655
OY 450 -AVQNCGGSIKTEEREIOCSGYTLPTKYAKKECCORC 485
Db 656 gasatcnalkpked-----akpcesqlc 678

RESULT 15
AAM81030
ID AAM81030 standard; Protein: 1496 AA.
XX
XX AAM81030;
XX
XX 10-MAY-1999 (first entry)
XX
XX DE Melanoma associated antigen MC50.
XX
XX KW MC50; melanoma gene-50; melanoma associated antigen; human;
XX T cell epitope; cancer; lung cancer; rhabdomyosarcoma; diagnosis;
XX therapy; vaccine.
XX
XX OS Homo sapiens.
XX
XX PN W09855133-A1.
XX
XX 10-DEC-1998.
XX
XX 04-JUN-1998; 98WO-US11533.
XX
XX 06-JUN-1997; 97US-0870941.
XX
XX (REGC ) UNIV CALIFORNIA.
XX (UWSC-) UNIV SOUTHERN CALIFORNIA.
XX
XX PI Deans RJ, Kan-Mitchell J, Minev BR, Mitchell MS;
XX
XX WPI: 1999-080820/07.
XX
XX N-PSDB: AAV99922.
XX
XX New MC50 melanoma associated antigen fragments - used to develop
XX products for the detection, treatment and prevention of
XX MC50-expressing cancers, e.g. melanoma, lung cancer or
XX rhabdomyosarcoma
XX
XX Claim 1; Page 45-49; 79pp; English.
XX
XX This polypeptide comprises a portion of a new human melanoma
XX associated antigen, designated MC50. The amino acid sequence was

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2001, 11:49:00 ; Search time 36.04 Seconds
(without alignments)
4346.536 Million cell updates/sec

Title: US-09-609-383-2
Perfect score: 6317
Sequence: 1 MVTGKAWVSEFLVLEVTSL.....QSGVVASLRPRVAQOPLIN 1184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SPTREMBL.16:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhmc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.sp._unclassified:*
14: sp.sp._vertebrate:*
15: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6317	100.0	1184	4 075339	075339 homo sapien
2	2898.5	45.9	599	6 019112	019112 sus scrofa
3	168.5	2.7	1948	4 013332	013332 homo sapien
4	164.5	2.6	1637	6 09XSV8	09XSV8 bos taurus
5	161.5	2.6	1077	4 09H324	09H324 homo sapien
6	155.5	2.5	7962	4 010465	010465 homo sapien
7	155	2.5	1496	4 092626	092626 homo sapien
8	154	2.4	779	5 097136	097136 manduca sex
9	154	2.4	7107	5 09V477	09V477 drosophila
10	152	2.4	1250	11 088971	088971 mus muscula
11	152	2.4	6658	5 076281	076281 drosophila
12	150	2.4	4123	4 075851	075851 homo sapien
13	147.5	2.3	788	5 022631	022631 caenorhabdi
14	147	2.3	1444	5 017591	017591 caenorhabdi
15	146	2.3	1239	5 061541	061541 drosophila
16	146	2.3	1239	5 09V3X0	09V3X0 drosophila
17	146	2.3	1302	5 061542	061542 drosophila
18	145.5	2.3	4488	5 09TXK4	09TXK4 caenorhabdi
19	144	2.3	1099	11 P97527	P97527 rattus norv

20	143.5	2.3	1034	11 035888	035888 rattus norv
21	143.5	2.3	4370	4 09H3V5	09H3V5 homo sapien
22	143	2.3	425	6 002661	002661 bos taurus
23	143	2.3	1360	5 09TYK4	09TYK4 caenorhabdi
24	142.5	2.3	1199	5 09V3G1	09V3G1 drosophila
25	142	2.2	1081	4 076065	076065 homo sapien
26	142	2.2	3570	4 099552	099552 homo sapien
27	141	2.2	852	4 09NS62	09NS62 homo sapien
28	141	2.2	1256	11 035158	035158 rattus norv
29	141	2.2	6642	5 001761	001761 caenorhabdi
30	140.5	2.2	432	4 09U0P1	09U0P1 homo sapien
31	140.5	2.2	2482	5 018263	018263 caenorhabdi
32	140.5	2.2	3375	5 09XTT5	09XTT5 caenorhabdi
33	139	2.2	1026	4 094780	094780 homo sapien
34	139	2.2	1100	4 094779	094779 homo sapien
35	139	2.2	2164	13 091AR9	091AR9 gallus gall
36	139	2.2	2222	5 097394	097394 drosophila
37	139	2.2	4162	13 098918	098918 gallus gall
38	139	2.2	4219	5 09NL87	09NL87 caenorhabdi
39	138.5	2.2	477	4 014887	014887 homo sapien
40	138	2.2	1599	5 009983	009983 caenorhabdi
41	138	2.2	2295	5 09XTD2	09XTD2 caenorhabdi
42	137	2.2	1083	11 09QW24	09QW24 rattus sp.
43	136.5	2.2	813	4 09ULX4	09ULX4 homo sapien
44	136.5	2.2	824	4 09UDY8	09UDY8 homo sapien
45	136.5	2.2	1527	5 09VZ24	09VZ24 drosophila

ALIGNMENTS

RESULT 1
075339 PRELIMINARY; PRT; 1184 AA.
ID 075339;
AC 075339;
DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
DE CARTILAGE INTERMEDIATE LAYER PROTEIN.
GN CLIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ARTICULAR CARTILAGE;
RX MEDLINE=98389785; PubMed=9722584;
RA Lorenzo P., Neame P., Sommarin Y., Heinegard D.;
RT "Cloning and deduced amino acid sequence of a novel cartilage protein (CLIP) identifies a proform including a nucleotide
pyrophosphohydrolase.";
RL J. Biol. Chem. 273:23469-23475(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Nakamura I., Okawa A., Ikegawa S., Takaoka K., Nakamura Y.;
RT "Genomic organization, mapping, and polymorphisms of the gene encoding
human cartilage intermediate layer protein (CLIP).";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Lorenzo P., Aman P., Sommarin Y., Heinegard D.;
RT "Pro-CLIP: Gene structure and chromosomal localization.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
CC EMBL: AF035408; AAC3838.1; -
DR EMBL: AB022430; BAA7662.1; -
DR EMBL: AF035455; AAF14689.1; -
DR EMBL: AF035448; AAF14689.1; JOINED.
DR EMBL: AF035449; AAF14689.1; JOINED.
DR EMBL: AF035451; AAF14689.1; JOINED.
DR EMBL: AF035453; AAF14689.1; JOINED.

DR InterPro: IPR000884; -
 DR InterPro: IPR001451; -
 DR InterPro: IPR002086; -
 DR InterPro: IPR003006; -
 DR Pfam: PF00047; 1g; 1.
 DR Pfam: PF00090; tsg.1; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN.1.
 DR PROSITE: PS00101; HEXAPEP_TRANSFERASES; UNKNOWN.1.
 DR SMART: SM00209; TSP1; 1.
 SO SEQUENCE 1184 AA; 132538 MW; 4449F0537CC99C3 CRC64;

Query Match 100.0%; Score 6317; DB 4; Length 1184;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 1184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGTKAWVFLEVLVETSVLGRQMTLQSVRYVPGKKNPSIFAKPADLTLESPEGNTLF 60
 1 MGTKAWVFLEVLVETSVLGRQMTLQSVRYVPGKKNPSIFAKPADLTLESPEGNTLF 60
 DB 1 MGTKAWVFLEVLVETSVLGRQMTLQSVRYVPGKKNPSIFAKPADLTLESPEGNTLF 60
 QY 61 NIDYPGKGYERLDAIRFYGGDRCARPLRLAARTTDMTPAGSTGVVHGSPEGEFWCL 120
 61 NIDYPGKGYERLDAIRFYGGDRCARPLRLAARTTDMTPAGSTGVVHGSPEGEFWCL 120
 DB 61 NIDYPGKGYERLDAIRFYGGDRCARPLRLAARTTDMTPAGSTGVVHGSPEGEFWCL 120
 QY 121 NREORPGONSNTRYRFLCPGSLRDRTERLWSPWSPKCSAACGOTGVOTRRLCLAE 180
 121 NREORPGONSNTRYRFLCPGSLRDRTERLWSPWSPKCSAACGOTGVOTRRLCLAE 180
 DB 121 NREORPGONSNTRYRFLCPGSLRDRTERLWSPWSPKCSAACGOTGVOTRRLCLAE 180
 QY 181 MVSICSEASEGQCHMGQDCTACDLTCPMGOVNADCDACMCDPMLHGAASLPGAPASG 240
 181 MVSICSEASEGQCHMGQDCTACDLTCPMGOVNADCDACMCDPMLHGAASLPGAPASG 240
 DB 181 MVSICSEASEGQCHMGQDCTACDLTCPMGOVNADCDACMCDPMLHGAASLPGAPASG 240
 QY 241 AATYLLTKTKLLTQDSDGRFRPLGICPBGKSLTKTKKFAPIYLTMPKTSIKAAITIK 300
 241 AATYLLTKTKLLTQDSDGRFRPLGICPBGKSLTKTKKFAPIYLTMPKTSIKAAITIK 300
 DB 241 AATYLLTKTKLLTQDSDGRFRPLGICPBGKSLTKTKKFAPIYLTMPKTSIKAAITIK 300
 QY 301 AEFVRAETPYMANNPETKARAGOSVSLCCGATGKPRPDYFWFHNHTLLDPSLYKHESK 360
 301 AEFVRAETPYMANNPETKARAGOSVSLCCGATGKPRPDYFWFHNHTLLDPSLYKHESK 360
 DB 301 AEFVRAETPYMANNPETKARAGOSVSLCCGATGKPRPDYFWFHNHTLLDPSLYKHESK 360
 QY 361 LVLRKLOOHAGEFYCKAOSDAGAVSKVAQLVITASDETPPCNVPFVPSYLRPLPHDFQON 420
 361 LVLRKLOOHAGEFYCKAOSDAGAVSKVAQLVITASDETPPCNVPFVPSYLRPLPHDFQON 420
 DB 361 LVLRKLOOHAGEFYCKAOSDAGAVSKVAQLVITASDETPPCNVPFVPSYLRPLPHDFQON 420
 QY 421 ATNSFYDVGRCPYKTCAGQDNGIRCRDAVONCCGISKTEEBEIOCSGTYLTPKVAKEC 480
 421 ATNSFYDVGRCPYKTCAGQDNGIRCRDAVONCCGISKTEEBEIOCSGTYLTPKVAKEC 480
 DB 421 ATNSFYDVGRCPYKTCAGQDNGIRCRDAVONCCGISKTEEBEIOCSGTYLTPKVAKEC 480
 QY 481 SCORCTETRSIVGRVSAADNGEPMRGHYMGNRSVMTGYKTFPLHVPDTERLVLT 540
 481 SCORCTETRSIVGRVSAADNGEPMRGHYMGNRSVMTGYKTFPLHVPDTERLVLT 540
 DB 481 SCORCTETRSIVGRVSAADNGEPMRGHYMGNRSVMTGYKTFPLHVPDTERLVLT 540
 QY 541 FVDRLQFVNTTKVLPENKGSVFEHFKMLRRKEPTITLAMEFNITPLGEVVGEDPMAE 600
 541 FVDRLQFVNTTKVLPENKGSVFEHFKMLRRKEPTITLAMEFNITPLGEVVGEDPMAE 600
 DB 541 FVDRLQFVNTTKVLPENKGSVFEHFKMLRRKEPTITLAMEFNITPLGEVVGEDPMAE 600
 QY 601 LEIPSRSFYQNGEPYIGKVKASVTFPLDPNISTATAQDNLFINDEGTFPLRTYGMF 660
 601 LEIPSRSFYQNGEPYIGKVKASVTFPLDPNISTATAQDNLFINDEGTFPLRTYGMF 660
 DB 601 LEIPSRSFYQNGEPYIGKVKASVTFPLDPNISTATAQDNLFINDEGTFPLRTYGMF 660
 QY 661 SVYFREDVTSEPLNAGVKVHLNSTOVKMPENHISTYKWLMSLNDPTGMEEGDFKFNOR 720
 661 SVYFREDVTSEPLNAGVKVHLNSTOVKMPENHISTYKWLMSLNDPTGMEEGDFKFNOR 720
 DB 661 SVYFREDVTSEPLNAGVKVHLNSTOVKMPENHISTYKWLMSLNDPTGMEEGDFKFNOR 720
 QY 721 RNKREDRTFLVGNLEIRERLFLNLDVPSRRCVKAAYSERFLPSEQIOGVIVISVNL 780
 721 RNKREDRTFLVGNLEIRERLFLNLDVPSRRCVKAAYSERFLPSEQIOGVIVISVNL 780
 DB 721 RNKREDRTFLVGNLEIRERLFLNLDVPSRRCVKAAYSERFLPSEQIOGVIVISVNL 780
 QY 781 EPTTGFSLNRRANGRDSVITGNGACVPAFCDDQSDAYSAVYLASLAGEELQAVESSP 840
 781 EPTTGFSLNRRANGRDSVITGNGACVPAFCDDQSDAYSAVYLASLAGEELQAVESSP 840
 DB 781 EPTTGFSLNRRANGRDSVITGNGACVPAFCDDQSDAYSAVYLASLAGEELQAVESSP 840
 QY 841 KFPANMAGVQPYLNLKLNRYRTDHEPDRVKKTAFOISMAMPNRSASENSGPIYAFENLR 900

DB 841 KFPANMAGVQPYLNLKLNRYRTDHEPDRVKKTAFOISMAMPNRSASENSGPIYAFENLR 900
 QY 901 ACEAPPSAAHFRFYQJEGDGYVNTVPEFNEDDPMSTEDYLAMWPMPEFRACVTKVKI 960
 901 ACEAPPSAAHFRFYQJEGDGYVNTVPEFNEDDPMSTEDYLAMWPMPEFRACVTKVKI 960
 DB 901 ACEAPPSAAHFRFYQJEGDGYVNTVPEFNEDDPMSTEDYLAMWPMPEFRACVTKVKI 960
 QY 961 VGPLEVNVSRNMGGTHTRTVGKLYGIRDVYSTREDQDPVNSAACLEFKSGMLYQDDRV 1020
 961 VGPLEVNVSRNMGGTHTRTVGKLYGIRDVYSTREDQDPVNSAACLEFKSGMLYQDDRV 1020
 DB 961 VGPLEVNVSRNMGGTHTRTVGKLYGIRDVYSTREDQDPVNSAACLEFKSGMLYQDDRV 1020
 QY 1021 DRTLKYIIPGSCRRASVNMLEHYLNLPLAVNNDTSEYTMALPLDPLGHNYGYITVT 1080
 1021 DRTLKYIIPGSCRRASVNMLEHYLNLPLAVNNDTSEYTMALPLDPLGHNYGYITVT 1080
 DB 1021 DRTLKYIIPGSCRRASVNMLEHYLNLPLAVNNDTSEYTMALPLDPLGHNYGYITVT 1080
 QY 1081 DDPRTAKEIALGRCFEDTSDGSSRIKSNVGVALTFCVYERGVGROSAPQYQSTPAQS 1140
 1081 DDPRTAKEIALGRCFEDTSDGSSRIKSNVGVALTFCVYERGVGROSAPQYQSTPAQS 1140
 DB 1081 DDPRTAKEIALGRCFEDTSDGSSRIKSNVGVALTFCVYERGVGROSAPQYQSTPAQS 1140
 QY 1141 PAAGTVGGRVPSRRORPASRGOROSGVASLRPPRYAQQPLIN 1184
 1141 PAAGTVGGRVPSRRORPASRGOROSGVASLRPPRYAQQPLIN 1184
 DB 1141 PAAGTVGGRVPSRRORPASRGOROSGVASLRPPRYAQQPLIN 1184

RESULT 2
 019112 PRELIMINARY; PRT; 599 AA.
 ID 019112
 AC 019112;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-NOV-1998 (TrEMBLrel. 05, last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, last annotation update)
 DE NUCLEOTIDE PYROPHOSPHOHYDROLASE (FRAGMENT).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95164702; PubMed=7860751;
 RA Masuda I., Hamada J., Haas A.L., Ryan L.M., McCarty D.J.;
 RT "A unique ectonucleotide pyrophosphohydrolase associated with porcine
 RT chondrocyte-derived vesicles."
 RL J. Clin. Invest. 95:699-704(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97473522; PubMed=9332376;
 RA Masuda I., Halligan B.D., Barbieri J.T., Haas A.L., Ryan L.M.,
 RA McCarty D.J.;
 RT "Molecular cloning and expression of a porcine chondrocyte nucleotide
 RT pyrophosphohydrolase."
 RL Gene 197:277-287(1997).
 DR EMBL; U83114; AAC48770.1; -
 KW Hydrolase.
 FT NON TER
 SO SEQUENCE 599 AA; 67432 MW; CCC4D08BCH1CFZD1 CRC64;

Query Match 45.9%; Score 2898.5; DB 6; Length 599;
 Best Local Similarity 91.1%; Pred. No. 5e-221;

Matches 544; Conservative 23; Mismatches 29; Indels 1; Gaps 1;

QY 588 PLCEVVGEDPMALLETIPSRFYQNGEPYIGKVKASVTFPLDPNISTATAQDNLFINDE 647
 588 PLCEVVGEDPMALLETIPSRFYQNGEPYIGKVKASVTFPLDPNISTATAQDNLFINDE 647
 DB 588 PLCEVVGEDPMALLETIPSRFYQNGEPYIGKVKASVTFPLDPNISTATAQDNLFINDE 647
 QY 648 EGDTPFLRTYGMFSVYFREDVTSEPLNAGVKVHLNSTOVKMPENHISTYKWLMSLNDPTG 707
 648 EGDTPFLRTYGMFSVYFREDVTSEPLNAGVKVHLNSTOVKMPENHISTYKWLMSLNDPTG 707
 DB 648 EGDTPFLRTYGMFSVYFREDVTSEPLNAGVKVHLNSTOVKMPENHISTYKWLMSLNDPTG 707
 QY 708 WEEDGDFKFNORRNKREDRTFLVGNLEIRERLFLNLDVPSRRCVKAAYSERFLPS 767
 708 WEEDGDFKFNORRNKREDRTFLVGNLEIRERLFLNLDVPSRRCVKAAYSERFLPS 767
 DB 708 WEEDGDFKFNORRNKREDRTFLVGNLEIRERLFLNLDVPSRRCVKAAYSERFLPS 767
 QY 124 WEEDGDFKFNORRNKREDRTFLVGNLEIRERLFLNLDVPSRRCVKAAYSERFLPS 183
 124 WEEDGDFKFNORRNKREDRTFLVGNLEIRERLFLNLDVPSRRCVKAAYSERFLPS 183
 DB 124 WEEDGDFKFNORRNKREDRTFLVGNLEIRERLFLNLDVPSRRCVKAAYSERFLPS 183

```

OY 768 EIOIGVIVINLEBRTGFLSNPRAMGRDSYITGPNACYPACFDQSDPADYSAYVLAS 827
DB 184 EIOIGVIVINLEBRAPSSNPRAMGRDSYITGPNACYPACFDQSDPADYSAYVLAS 243
OY 828 LAGELEAVESSPKFNPAIGVOPYLKLNLRTHDEDPVRYKTAFOISMAKPPNSAE 887
DB 244 LAGELEAVESSPKFNPAIGVOPYLKLNLRTHDEDPVRYKTAFOISMAKPPNSAE 303
OY 888 ESNGITVAFENLRACEAPPSAHRFRYOIEBDRDYNTVTPNEDDPMSWTEDYLAAMPK 947
DB 304 ESNGITVAFENLRACEAPPSAHRFRYOIEBDRDYNTVTPNEDDPMSWTEDYLAAMPK 363
OY 948 PHEPACITKIVGPLEVNVASRMWGGHRTVSKLGIRVRSRTRPDOPNSAACIE 1007
DB 364 PHEPACITKIVGPLEVNVASRMWGGHRTVSKLGIRVRSRTRPDOPNSAACIE 423
OY 1008 FKCSGMLYDODRVDRTLVKVIPOGSCRASVNPMLHEYLNNLPLAVNNDTSEYTMALPL 1067
DB 424 FKCSGMLYDODRVDRTLVKVIPOGSCRASVNPMLHEYLNNLPLAVNNDTSEYTMALPL 483
OY 1068 DPLGNHYGTYITDODDPTAKFIALGRCFEDGSDSSRLMKSNVVALTFNCVEYQVGRQ 1127
DB 484 DPLGNHYGTYITDODDPTAKFIALGRCFEDGSDSSRLMKSNVVALTFNCVEYQVGRQ 543
OY 1128 SAFQYLOSTPAOSPAAGTVGGRVPSRROORASRGOGOSGVVASLRFPVAAOPLIN 1184
DB 544 SAFQYLOSTPAOSPAAGTVGGRVPSRROORASRGOGOSGVVASLRFPVAAOPLIN 599

RESULT 3
O13332 PRELIMINARY: PRT: 1948 AA.
ID O13332
AC O13332: Q15718; Q16341;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE, RECEPTOR-TYPE, S PRECURSOR (EC 3.1.3.48)
DE (PROTEIN-TYROSINE PHOSPHATASE SIGMA) (R-PTP-SIGMA) (PTPRS)..
GN PTPRS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FETAL BRAIN;
RX MEDLINE=96102179; PubMed=8524829;
RA Pulido R., Serra-Pages C., Tang M., Streuli M.;
RA "The LAR/PTP delta/PTP sigma subfamily of transmembrane protein-
RT tyrosine-phosphatases: multiple human LAR, PTP delta, and ptp sigma
RT isoforms are expressed in a tissue-specific manner and associate with
RT the LAR-interacting protein Lip-1."
RL Proc. Natl. Acad. Sci. U.S.A. 92:11686-11690(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96255038; PubMed=8992885;
RA Ende N., Rutledge S.J., Opas E.E., Vogel R., Rodan G.A., Schmidt A.;
RT "Human protein tyrosine phosphatase-sigma: alternative splicing and
RT inhibition by bisphosphonates."
RL J. Bone Miner. Res. 11:535-543(1996).
RN [3]
RP SEQUENCE OF 1503-1589 FROM N.A.
RA lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA Dangnan L., Eiler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attix C., Andreise T., Frankheim M., Amico-Keller G., Cosfield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA Arellano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrino A.V.;
RT "Sequence analysis of a 2.5 Mb region in 19p13.3."
RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 1503-1589 FROM N.A.
RX MEDLINE=92119637; PubMed=1370651;
RA Adachi M., Sekiya M., Arimura Y., Takekawa M., Itoh F., Hinoda Y.,
RA Imai K., Yachi A.;
RT "Protein-tyrosine phosphatase expression in pre-B cell NALM-6."
RL Cancer Res. 52:737-740(1992).
CC -1- FUNCTION: INTERACTS WITH LAR-INTERACTING PROTEIN LIP-1.
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O = PROTEIN
CC TYROSINE + ORTHOPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE FOUND IN DIFFERENT
CC TISSUES DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: DETECTED IN ALL OF THE TISSUE TESTED EXCEPT
CC FOR PLACENTA AND LIVER.
CC -1- SIMILARITY: EXTRACELLULAR REGION TYPIC OF A CAM FAMILY (3 IG-LIKE
CC DOMAINS FOLLOWED BY 8 FIBRONECTIN TYPE III-LIKE DOMAINS), AND A
CC CYTOPLASMIC REGION COMPOSED OF TWO PTPASE DOMAINS.
DR EMBL: U35234; AAC50299.1; -.
DR EMBL: U40317; AAC50567.1; -.
DR EMBL: AC005790; AAC62832.1; -.
DR EMBL: S78080; AB21146.1; -.
DR HSSP: P18052; 1YFO.
DR MIM: 601576; -.
DR InterPro: IPR000242; -.
DR InterPro: IPR000387; -.
DR InterPro: IPR001777; -.
DR InterPro: IPR003006; -.
DR Pfam: PF00041; In3; 8.
DR Pfam: PF00047; Ig; 3.
DR Pfam: PF0102; Y_phosphatase; 2.
DR PRINTS: PR00104; FNTYPEIII.
DR PRINTS: PR00700; PRTYPHPTASE.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; FALSE_NEG.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 2.
DR PROSITE: PS50055; TYR_PHOSPHATASE_ptp; 2.
DR SMART: SM00194; PTPc; 1.
DR Hydrolase: Receptor: Glycoprotein: Signal: Transmembrane:
KW Cell adhesion; Immunoglobulin domain; Duplication;
KW Alternative splicing; Repeat.
FT SIGNAL 1 29
FT CHAIN 30 1948
FT DOMAIN 30 1282
FT TRANSMEM 1283 1303
FT DOMAIN 1304 1948
FT DOMAIN 32 124
FT DOMAIN 127 234
FT DOMAIN 241 327
FT DOMAIN 329 423
FT DOMAIN 525 615
FT DOMAIN 618 717
FT DOMAIN 720 831
FT DOMAIN 834 926
FT DOMAIN 928 1033
FT DOMAIN 1036 1151
FT DOMAIN 1411 1657
FT DOMAIN 1700 1948
FT DOMAIN 641 644
FT DISULFID 54 107
FT DISULFID 156 216
FT DISULFID 266 311
FT ACT_SITE 1589 1589
FT ACT_SITE 1880 1880
FT CARBOHYD 263 263
FT CARBOHYD 308 308
FT CARBOHYD 733 733
FT CARBOHYD 940 940
FT VARSPPLIC 190 198
FT VARSPPLIC 236 239
FT VARSPPLIC 617 1034
FT VARSPPLIC 784 792
FT VARSPPLIC 1035 1035

```

FT	VARSPPLIC	1350	1365	MISING (IN PTPS-MED).
FT	VARSPPLIC	1366	1366	S -> G (IN PTPS-MED).
FT	CONFLICT	309	310	YT -> YHP (IN REF. 2).
FT	CONFLICT	428	429	SA -> RP (IN REF. 2).
FT	CONFLICT	742	745	LGVP -> RSPA (IN REF. 2).
FT	CONFLICT	765	773	GABRGPR -> RREARGRRS (IN REF. 2).
FT	CONFLICT	910	910	R -> P (IN REF. 2).
FT	CONFLICT	986	995	AAEAGGNAV -> GRISRRRTL (IN REF. 2).
FT	CONFLICT	1195	1196	TV -> SL (IN REF. 2).
FT	CONFLICT	1431	1431	F -> S (IN REF. 2).
FT	CONFLICT	1546	1546	E -> D (IN REF. 4).
FT	CONFLICT	1587	1587	V -> A (IN REF. 4).
FT	CONFLICT	1705	1705	N -> K (IN REF. 2).
SO	SEQUENCE	1948 AA:	217080 MM:	7DC049EC03171136 CRC64:

Query Match	2.7%;	Score 168.5;	DB 4;	Length 1948;
Best Local Similarity	20.2%;	Pred. No. 0.0006;		
Matches 195;	Conservative 117;	Mismatches 366;	Indels 289;	Gaps 48;

QY	295	KAATIKKEEVAEPPYVWMPDETARAGOSVSSJCCATKCPRPDXFWVHNDTLDPST	354
Db	121	KLTVLRDQLPSSGPNIDMGRLKYVERTTATMLCAASNPPPE--ITMFKDLPDPSSA	179
QY	355	YKHESK-----LVLRKLOHQAGEYFCCKAOSDGAAYSKVAOLITASDE	399
Db	180	SNGRIKOLRSETFESTPIRGALQJESSEETDQCKEYCATNSAGVHYSSPANLYARELRE	239
QY	400	TPCRNPVESTILRLPHCFQNA--TNSYIYVGR--CP--YKTCAGQDNGIRCDVAONCCG	456
Db	240	VR--RVARRSTILPESHHEIMDGNVNTTCVAVGSPMPYKMGQAED-----	284
QY	457	ISKTEEREIOCSGTYLTPEYKAECCSCORCTETRSIYGRVSADNGBPMHFGHYVGNRS	516
Db	285	--LTPEDDMGVGRVNLDELTVKOSANTCYAMSSL--GVLEAV-----AQ	325
QY	517	VSMTGY--KGFTYLHPDQTERLVLTVEYDLQKFTVNTKYLPEFKKGSAYFHEIKMLRKE	575
Db	326	ITVKSILKPAKGTPEVNTENTATISLITWDSGNDPVPASVYLEYKSKGODPQOLK-----	379
QY	576	PITLEAMETNIPRGVEVGEDPMALLET-----PSSSFYKONGEPITGKVKAS	623
Db	380	----EDITTTTYSIG--GSPNSEETIYWSAVNSIGQGPSSSVYTRGE--OAPS	428
QY	624	VTELDPNRISTATAQTDLNFINDGDTPELRTYGFMSVDFREBVSSEPLNAGKYVHND	683
Db	429	A----PRNVOQARMLSATNTIYQMBE--PVEENGLIR--GYRYIYME-----	467
QY	684	STOVKAPHEHISTYKVLMSLNDPTGLMBEEDGFKENORRKNKREDRTLVGNLEIRERLEN	743
Db	468	----PEH-----PVGNW-----QKHVNDLSLITTVGSL-----	491
QY	744	LDVESHRCFVKRAYSERFLP--SRQI-----QGVVISTINL--BPRGFSLNPRANG--	794
Db	492	----LEDETYIVRALFTSVSDGRLSDPTQVKTQGVGPGPMNLRAARKE--TSITLSWSP	547
QY	795	--REDSVI-----TGNBACVPAFCDDQSPDAYSAVYASLASGEELOQVSSPKENPNA	846
Db	548	PROGSLIKETLPEFGHGREY-----GRIFDPTTSYVEDLKPNEYAFRLAAR--SPQG	601
QY	847	IGVPOPLINKLNTRRIDHEDPR-----YKTAFOJISMAKPRENSAEESNGPIYAE--	897
Db	602	LGAFETPVVORTLOSKSPAPRODVCKVSVYSTALIVMSRPPP--ETHNGALGVSVAX	658
QY	898	-----NLRACEEAPPSAHRFQOIE--GDRYDYNTPFNDDP-----MSWTD	940
Db	659	RPLGSEDPKEKEYNGLPPTTQILLLEALEKWTQYRTTYVAHTEVGGRESSPVVAKTDD	718
QY	941	YLAAMPKEME--FRACYIKVKIVGDEVNVRNNGGTHRTATYGLKYGIRDP--STRD	995
Db	719	VPSAPPRKXVBAELNMTALINVLMLGVP-----GRHQGQLGR--YQYHYHRRMGAGG	768
QY	996	RDDPNVSACLERCSGMLYD--ODRVDTLVKVIPOGSCRRASVNBMLHEYLVNHLPLAV	1054

Db	769	RGPRINDV-----MLDADQEND-----	787
QY	1055	NNDISETMLAPLEDPLGHANIGITVYDQDPRIKAFETALRCEDGTSOGSSRIMKSNVGA	1114
Db	788	--DPAEEEMV-----ITNLOPETAVSITVA-AYTMKODGA-----RSKPRV	826
QY	1115	LTFNCVEROVROSGAFQYLOSTPAOS-----PAAGTVGGRVPSRPOOASNGGROSGV	1168
Db	827	VTKAAV--LGRPTL--SYQOTPEGSLMARMEPPATAEDQYLGIRLQ-----FGREDSTP	877
QY	1169	VASLRFP	1175
Db	878	LATLEFP	884

RESULT	4			
09XSV8				
ID	09XSV8	PRELIMINARY:	PRT:	1637 AA.
AC	09XSV8			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		
DT	01-NOV-1999	(TREMBLrel. 12, last sequence update)		
DT	01-MAR-2001	(TREMBLrel. 16, last annotation update)		
DE	SCO-SPONDIN	(FRAGMENT).		
GN	SCO-SPONDIN			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_Taxid:9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RT	Gobron S., Creveaux I., Didier R., Weiniel R.,			
RA	"Characterization of cattle SCO-SPONDIN."			
RL	Submitted (MAR-1993) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AJ134488; CAB46239.1; ..			
DR	HSSP; P56682; ICCV.			
DR	InterPro; IPR000359; ..			
DR	InterPro; IPR000884; ..			
DR	InterPro; IPR001007; ..			
DR	InterPro; IPR001545; ..			
DR	InterPro; IPR002919; ..			
DR	Pfam; PF00007; Cys_knot; 1.			
DR	Pfam; PF00090; tsp_1; 11.			
DR	Pfam; PF01826; TIL; 5.			
DR	PROSITE; PS01225; CTCK_2; 1.			
DR	PROSITE; PS00261; GLYCO_HORMONE_BETA_1; UNKNOWN_2.			
DR	PROSITE; PS01208; WVEC; UNKNOWN_1.			
DR	SMART; SM00041; CT; 1.			
FT	NON_TER			
SO	SEQUENCE	1637 AA; 173346 MW; 4CSBAB1D346C925 CRC64;		

Query Match	2.6%;	Score 164.5;	DB 6;	Length 1637;
Best Local Similarity	21.6%;	Pred. No. 0.00095;		
Matches 100; Conservative	26;	Mismatches 160;	Indels 177;	Gaps 22;

QY	94	ARTIDMNP	----	AGSTGOVYH	-GSPRRGFC-	LNRRQRRQGNCSNTVAFLC	-----	133
		::: :::						
Db	646	SRMSAMSPCSRSCG	PPAQQSRFRSTSTSGSVAPECREEDQSOSPDPGSPCPPLCLQGTGR					705
QY	140	-----	PGSLRRDTE	-----	RIWSPWSPMSKCAAGCGQGVORRTICL			178
		::: :::	::: :::		::: :::			
Db	706	SLGDSWLQDGGQCSCT	PEGLICEDACAGLANTPMSPSDCPVSCG	-G6N0YRTRVCY				764
QY	179	AEMVSLCSEASEEGCHCMGD	-----	CTACDLTTC	-PMGVNADCDACMCDFM			225
Db	765	A-----	SAPPRGGSPCLTGPDVQSQRCGLWPCCALPDTCSMGPMGPCSRSCG	-----				810
QY	226	LHGAVSLPGCAPAGAAIYLLTK	-----	TRKLLTQTSDDGRFRIRGLCLCPGSKILKI				277
		::: :::		::: :::				
Db	811	-----	PGLASRASQDCLLAEEAPACNSTSPILDTQA	-----	CTAG			846


```

Db 2688 GPNITTSKGC---NIVTEKTCILEILN---STRDAGGY-----SCIEINERAGR----- 2731
QY 216 CDACMCQDFMLHGAVSLFGAPASAAIYLLTKTPKLLTQDSDGRFRIPGLCPDGKSL 275
Db 2732 -DVC-----GALVSTL----- 2741
QY 276 KITVKKAPVILTMPTKSLKATTKAEFVRAETPYVMNPETKARRAGOSVSLCKATKG 335
Db 2742 -----EPPYFTELEPLAANGDSVSLCCQVAGT 2770
QY 336 PRDQKTYWYHNDLLDS-----LYHESKLYLRKLOOHAGEYFCKAOSDAGAVSKV 389
Db 2771 PE-ITVSWYMGDKRLRPTPEYRRTFTNNVATLVENKYNINDSGEYTCKAENISGTASKT 2829
QY 390 AOLI-----VTASDETPCNVPP-----ESYLRLPHDCQ- 419
Db 2830 VFRIOERQLRPPSFARQKLDIEQVGLFVTLTCLRLNGSAPLQVCWYRGVLLR-DHENLQT 2888
QY 420 -----MATN-----SFYDVGRCPYKTC 437
Db 2889 SFVDNVAATLKILQTDLSHSGQVSCSASNPLGTASSARLAREPKKSPFDIKPVSIDVT 2948
QY 438 AGOODNGIRCRDAVONCCGIS-KTEREIOCGS-YTLR-----TKARECSQ-R 484
Db 2949 AGESAD-FECHVTGAOPMRITWSKDNKEIRPGGNITTCVGNTPHLRIKLVKGDGSGQYT 3007
QY 485 CTETRSIVGRVSA-----ADNGEPMRPHVWNGSRVMTGKGT 525
Db 3008 CQATNDVGCMDKSAQSLVKEPEKPFVKLASAKYAKOGESITQLECKTISGPEIKVSWFRND 3067
QY 526 FLIHPQDTERLYLTFVDRLOKRVNTKVLPPNK----- 559
Db 3068 SELHESWKYN-----MSFINVALLITINEASAEDSGYICAHNGVGDASGTL 3117
QY 560 --KGSAPVEHEIKMLRKKEPTILEAME-TNITPLGEVVGDPMALEIPRSFTRONGEPY 616
Db 3118 TVKAPPV-----TOKSPVGAUKGSDVILQCEISGTTPPEVWVWKRKQVR-NSKRF 3169
QY 617 IGKVAASVTEFLDPNISTATAQTDLNFDNG-DTF-----PLRTYGMFS----- 661
Db 3170 KITSKHPTNLHILMLASDVEGHCATNEGSDTSCSVKFKPPRVKLSLSTLI 3229
QY 662 ---VDFREDEVS-EPLNA-----GKVYHLDSTOVKMPENHISTVKLMS----- 700
Db 3230 GDAVELRAIVGEGFOPISVWMLKDEGEVIRESENTRISFIDNATLQLSPEASNSGKYIC 3289
QY 701 -LNPDTGLMEEGDEKPFENOR--NKREDREFVGN----- 733
Db 3290 QIKNDAGMRCSAVILVLEPARITIEKPPMTVTGPNPALECVYTGPELSAKWFKDGRE 3349
QY 734 -----LEIRERRLFNLDVPS---RRCFVKVRAVRSERFLPS- 767
Db 3350 LSAOSKHITFINKVASLIKPCAMSDKGLAFEVKNSVGSNCTVSYHV--SDRIVPPS 3407
QY 768 -----EDIOGVIVIVINLEPRTGFLSNPRANGF--DSVITGPNACVPAFC- 813
Db 3408 FIRKLKQVNAVLGVSYLECRVSG-SAPISYGMFQDNEIYSGPK--CQSSFSNACTLN 3464
QY 814 ----DQSPDAYSAVILASLAGEE---LQAVESPKF--NPMAGV 849
Db 3465 LSLLEPDTGTIYCVANANVAGSDCSAVLTVOEPSPFQDPDSEV 3510

```

```

RESULT 7
ID 092626 PRELIMINARY: PRT: 1496 AA.
AC 092626:
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, last sequence update)
DE 01-MAR-2001 (TREMblrel. 16, last annotation update)
GN WHELOBLAST KIAA0230 (FRAGMENT).
KIAA0230.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RX MEDLINE=95048383; PubMed=7959781;
RA Weiler S.R., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,
RA Trent J.M.;
RT "Assignment of a human melanoma associated gene M650 (D2S448) to
RT chromosome 2p25.3 by fluorescence in situ hybridization.";
RL Genomics 22:243-244(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;
RT "Identification of a novel melanoma gene (M650) - likely the gene for
RT IL-1 receptor antagonist - which encodes epitopes recognized by human
RT cytolytic T lymphocytes.";
RL Submitted (Oct-1999) to the EMBL/Genbank/DBJ databases.
CC -1 SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL: D86983; BAA13219.1;
DR EMBL: AF200348; AAF06354.1;
DR HSSP: P05164; ICXP.
DR InterPro: IPR000483;
DR InterPro: IPR001007;
DR InterPro: IPR001536;
DR InterPro: IPR001611;
DR InterPro: IPR002007;
DR InterPro: IPR002016;
DR InterPro: IPR003006;
DR Pfam: PF00047; Ig; 4.
DR Pfam: PF00093; vwc; 1.
DR Pfam: PF00141; peroxidase; 1.
DR Pfam: PF00560; LRR; 5.
DR Pfam: PF01463; LRRT; 1.
DR PRINTS: PR00457; ANPEROXIDASE.
DR PROSITE: PS01208; WPCF; UNKNOWN_1.
DR SMART: SM00214; vwc; 1.
DR NON_TER 1
FT SEQUENCE 1496 AA; 167209 MW; E9B9A706BFB1ABFF CRC64;
SQ

```

Query Match 2.5%; Score 155; DB 4; Length 1496;
Best Local Similarity 18.5%; Pred. No. 0.0047;

Matches 186; Conservative 115; Mismatches 352; Indels 352; Gaps 46;

```

QY 297 ATIAEFVRAETPYVMNPETKARRAGOSVSLCKATGKRPDKYFYHND----- 347
Db 251 AITPEELNCEPRITSEPDADVTSGNTYVETCRAGGNKPE-IWLRNNNELSMKTD 309
QY 348 --TLDPSTLYKHESKLYLRKLOOHAGEYFCKAOSDAGAVSKVAVOILVYASDETP- 401
Db 310 RLNLDD-----DGTLMIQWTOETDGTITQCMKKNVAGEVKTQEVTLRFEGSPARPTFY 363
QY 402 -----C-----NFPESYLIR-----LPHDCFQMATNSFYVDGRCPVKT 436
Db 364 QPONTFVLGEVSVTLCSATGHPHPRIISMTRGDRTPLPDPKYNITPS-----GLYIGN 418
QY 437 CAGQODNGIRCRDAVONCCGISKT-----EER-----ELQCSGYTLP 473

```

```

Db 419 WV-QDSEGYACATNNIDSNHATFIIYQALPQFVTVPODRVIEGQVDFQCEAKGNP 477
QY 474 -----TKVAKESGOCRTFTSRIVGRVSA---DNCEMRGHHVYMGSRVSMGCGK 524
Db 478 PVIANTKGSQSLVDRLHVLSSGTRISGVALHDOGOECAVNITIGOKV----- 530
QY 525 TETLH-----VPDTERLVLFVDBLQKFVNTTKV----- 554
Db 531 --VAHLTVOPRTVPFASIPSDT-----IVEGANVQLPCSSOGEPERA 572
QY 555 LPPNKGK-----SAVH-----ETIKMRKRPITLMEATNIIPLGEVGEDPMMAELE 602
Db 573 ITNNKGVQVTESEKFIHISPEGLTINDVPADAGRECVARNTIGSASV---SMVLSVN 629
QY 603 IPRSPFYROWGEPIYIG-KVYASVTFIDPRNISTATAQDNLNINDEGDFPLRTGMFS 661
Db 630 VPPVS---KRGDPFVATSTVEALATVDRAINSTRT-----HLFSPRSPDILLALFR 679
QY 662 VDFREYTSERPLNAGVKVHLIDSTQVKMPREHISTVKLMSINPTGLMEEGDEKFEFEN--- 718
Db 680 YP-RDPTVYQARAGEI---FERTLQIIGHVGHGLMVDLNGTSYHNDLVSPQYLNLIA 735
QY 719 -----QRRKREDRTF-----LVGNIE----- 735
Db 736 NLSGCTAHRREVNNCSMCFHOKYRTHDGTGNNLQHPMMWGLTAFERLTKSYENGENTP 795
QY 736 --FRRERLEN---LDVPESHRCFVKYRAYRSEERFLPSEOIGVIVISINLEPRTGFLSNP 790
Db 796 RGINPHRLIGHALPMPR---LVSTTLIGTEVTVPDQETHMLMO----- 837
QY 791 RANGRF-----DSVITGPNACVPAFCDDQ-----SPDAYSAYV---LASLAGEE 832
Db 838 --WGFELDHLDSTVVALSGA---RFSGQHCNSVCSNDPCCSVMLPRPDSARARSARC 892
QY 833 LQAVESSPKFNPAIGV-----PQRYLKNL-----YRRTDEDDPVK----- 870
Db 893 MFEVRSSPVGSGMTSLMSNVYPREQINQLTSTYIDASNYVYGSTEHARSIRDLASHRGL 952
QY 871 -----KTAFQISMAKPRPNSAESNGPIYAF---EMLRACSEAPPAAPFRFQIE 918
Db 953 LRQGIORSKRPLLPRATGPRTECMNDENSPICFLAGHANEODIGLTMHTLMFR-E 1011
QY 919 GDRYDNTVAFNEDPMSTEDYLAWMPKMEPRACYIKVIVGPLEVNVSRN-----M 973
Db 1012 HNRITATELKNP---HMDGDTI-----YETKIKYGAELQHTTYGHMLPKIL 1056
QY 974 GGTHTRTVGLYIGRIDVSTRDRDQPNVSAACEFKCSGMLYDQDRVDTLVVITPDGSC 1033
Db 1057 GEVGMRTLGEYHGY-----DPOINGIFNAFATAF---RGNHTL----- 1093
QY 1034 RRSVPMMLHEVYNHLPLAVNNDTSEYTMAP-----LDPL 1070
Db 1094 ----VNPLLYRLDENFOPIAQDHLPLHKAFFSFRIVNEGIDPL 1134

```

RESULT 8

```

ID 097136 PRELIMINARY: PRT: 779 AA.
AC 097136;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, last annotation update)
OS FASCIICLIN II GPI-LINKED ISOFORM.
OC Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Dityrsea;
OC Sphingioidea; Sphingidae; Sphinginae; Manduca.
OX NCBI_Taxid=7130;
RN [1]
RP SEQUENCE FROM N.A.
RA Wright J.W., Snyder M.A., Combes S., Copenhaver P.F.;
RT Fasciclin II and neuronal migration.
Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

```

```

CC -1 SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL: AF103899; AAD17918.1; -.
DR InterPro: IPR001777; -.
DR InterPro: IPR003006; -.
DR InterPro: IPR003598; -.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00047; fn3; 5.
DR SMART: SM00408; IGC2; 1.
SQ SEQUENCE 779 AA; 86493 MW; BD9600E4DE02BC5 CRC64;

```

Query Match 2.4%; Score 154; DB 5; Length 779;
 Best local similarity 20.3%; Pred. No. 0.0021;
 Matches 132; Conservative 88; Mismatches 203; Indels 228; Gaps 35;

```

QY 294 LKAATTKAEFVRAETPYVMNPETKAR-----RAQSVSLCKATGKPRPDYFYVHND 347
Db 211 LABRNKILEVYTA-----PEMERESRVEIKEGESAATCKARKGP-PPRYTWIKAS 261
QY 348 TLLDP-----SLYKHSKLVLLKLOHQAGEFCKAQSDAG-----AVKSKVAQ 391
Db 262 TRENLATTSRFVSNEISGLTFDVRAGDYGYKICSAVNNAGONETEIEVEVLVKKRIFE 321
QY 392 LIYASDET-----PC-----NPVESYLRLPH-DEFOANTSEFYDYVGRCPVTCAGQ 441
Db 322 LKNTTAPQDQEGRLCKAIGRAPRISFKLSNDRFLNPN---DGRITTIETSSROT 377
QY 442 DNGIRCDAYONCGISKTEEREIQSGYTLPTKVAKESGOCRTETRSIVRGVSAADN 501
Db 378 GDQMESTGVI-TISTLNRTDGLYEC-----VAENDG 408
QY 502 GEPMRGCHV-----YMGNSRVSMTGKGTFTLHVPOTERLVLFVDR 544
Db 409 GEARRRGHLVVEKRPSEHMPVPIWANNQPNLS---CIAESIPINATIKKRFRELD 464
QY 545 LQ-----KPYNTT-----KVLPPKKGSAVHEIKMLRKRPITLLEAM 582
Db 465 VETHMLKIRSGPISITITTTPLDQDLFGYKCIATNTHGAE-HILOFRAYRPGAVVOA 523
QY 583 ETNIIPLGEV---VGEDPMAELEIPSRSP---YRONGEPIYKVASVTFIDPRNIS- 633
Db 524 KOELITATSVTFQIG--PAEMGPRILATYAQYKNGN-----FDMWLAQN 568
QY 634 -TATAQDNLNINDEGDFPLRTYGMFSYDF----- 664
Db 569 RTWSANSNSLYVEN-----LRP-MFTYDFEFAANQVGVAMGSPLTIVIMPRSPPE 620
QY 665 ----RDVTSERPLNAGKVKVLIDSTQVKMPREHISTVKLMSINPTGLMEEGDEKFEFEN--- 711
Db 621 QPKRREDISESLIHGN--YADRYELK-----KAVPRAHNBEPDLYETQYCPVL 667
QY 712 ---GDFKFNORRNKRKD-RTPLVGNLEIRERRLFNLDPESHRCFVKYRAYRSEERFLP 766
Db 668 KVSQDMRVADESLCYVEKQIESFEVINVEYRGLK-----PDT-RYMHMLRAHNVLFESLP 720
QY 767 SEQIOGVVISINLEPRTGFLSN--PRANGRD---SVITGPNACVPAFC 812
Db 721 AQLYVOTALGEYNSE-----SNEVPROPFGFYDVTAASEALKPSSAITTTC 765

```

RESULT 9

```

ID 09VAF7 PRELIMINARY: PRT: 7107 AA.
AC 09VAF7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, last annotation update)
OS BT GENE PRODUCT.
GN BT OR CG1479.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```


OC Ephydroidea; Drosophilidae; Drosophila.
 RX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champé M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jattal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paule J.M.,
 RA Palatzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shier B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -I- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -I- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
 CC DOMAIN.
 DR EMBL: AE003843; AAF59316.2; -;
 DR HSSP: P40189; 180U
 DR Flybase; FBgn005666; bt.
 DR InterPro; IPR000255; -;
 DR InterPro; IPR000577; -;
 DR InterPro; IPR000719; -;
 DR InterPro; IPR001777; -;
 DR InterPro; IPR002290; -;
 DR InterPro; IPR003006; -;
 DR Pfam; PF00041; fn3; 39.
 DR Pfam; PF00047; 19; 20.
 DR Pfam; PF00069; pkinase; 1.
 DR PRINTS: PR00014; ENTPEITL.
 DR PROSITE; PS00933; FG_Y KINASES.1; UNKNOWN.1.
 DR PROSITE; PS00102; PHOSPHOPANTHETHEINE; UNKNOWN.1.
 DR PROSITE; PS0107; PROTEIN_KINASE_ATP; UNKNOWN.1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KM ATP-binding; Repeat: Serine/threonine-protein kinase; Transferase.
 SO SEQUENCE 7107 AA; 794479 MW; CC852ABC29ADBFRCRC64;

Query Match 2.4%; Score 154; DB 5; Length 7107;
 Best Local Similarity 19.3%; Pred. No. 0.061;
 Matches 160; Conservative 101; Mismatches 257; Indels 310; Gaps 39;

QY 234 GGAPASGAIA-----VLTTPKLLTQTD-----SDGRFRIPGLCPDGK---- 272
 Db 2165 GGASITGYIVERKDPNTGKMQKALETSTPDCARVNDLIAGNKYQFRIMAVNKGAKSKPS 2224
 QY 273 --SIKITKVFAPVLTMPK---TSLKATIKAEFAVETPVVMMNPETARRAGOSVS 327
 Db 2225 EPSDQMTKKDRFAP---PRIDRININDITIK-----AGGHIR 2258
 QY 328 LCKKATKPRDKYFWYNDITL--DPSLY-----KHESKLVLRKLOOHQGEYFAQOS 380
 Db 2259 FDIKVSQEP-FATKWLHLNKAERLNDSDNIMDESYTKTLTYISKRFHSGKTLTKAEN 2317
 QY 381 DAGAVKSVVAOLITVASDETPCNPVPESSYLIRLPDHFQNAATNSFYDVGR--CPVKTC 438
 Db 2318 ESGRDEASFETVILD-----KCPPEGPPLRVLT-----DVHREGCKLKWNA 2357
 QY 439 GQDDGICRCRAVONCCGISKTEEREIQCSTYLLPTKAKESQO-KQTERISIVRGRVS 497
 Db 2358 PLDDGGLPIIDHYI-----TEKMDVE-SGRMLPSGRFKESFALNNLEPSHEYKFRVL 2408
 QY 498 AAD---NGEPMRFGHYVGNRSVSWTGYKGTFTLVPODTERLVLTVEYDRLOKEVNTTKV 554
 Db 2409 AVNTEGESEPLEGEGSVLAKNPFDGPKRGF-----PE-----AVDMKDHVDLWVR 2455
 QY 555 LPFKKGSVAVHEITMLRKPEPT---LEAMETNIIPLGE-----VGEDPMAELEI- 603
 Db 2456 PPIINDGSPIDIGYV--VERREKGTDMKIKGTETIIPICGECCKATVPTLINCCEYERVK 2513
 QY 604 -----PSKSEFRONGEPIYIGKVASVTELPDRNISTATAQTDLNF--INDEGDTF 652
 Db 2514 AINAGPGEPSDA-----SKPIITPKRLAKRIDKNIRITNFKSGEPIFLDINISGPA 2568
 QY 653 PLRT-----YMFVSVERDE 667
 Db 2569 PVTWNQNNKSVQTSFSHIENLPYNTKYINNPERKDTGLYKISAHNFYQDDOVEEIN 2628
 QY 668 VTSEPLNNGK-----VKVHLDSTQVAKP-----EHISVKLMSINPTGLWEEGD 713
 Db 2629 IITKP--GKEPGLEVSEVHDGCKLWKPKPKDGDGPEVSELYEAKDPDGTGLMLPYG- 2684
 QY 714 FFEQORRRKRDRFVLGNLEIRERRFLNDVPSRRCFVKVRAVRSERFLPSEOTIGV 773
 Db 2685 -----RSDGPE-----YNDGLVPGHDYK-F 2704
 QY 774 VISYINLEPRIGFLSNPRAMGFDSVITGPNACYPACDDOSPATSAIYVLASIGEL 833
 Db 2705 RYKAVNKE-----GESEPLET-----LGSIIADP 2729
 QY 834 GAVESSPKFNPAIGVPOPYLNLKLVYRRTHDEDPVKKKTAFQISMARPPNSAEESNGPI 893
 Db 2730 FSVPTKP-----GVPEP-----TDW-----TANKVELAMPEP--ASDGSPI 2764
 QY 894 --YAEFNT-----RACEAPSAAHFRFYQIEGDRDYNTVPNE 931
 Db 2765 QGYIVEVKNQKYSPLWEKALETNSPTPTATVQGLIGNEGYQFRVALNK 2812
 RESULT 10
 ID 088971 PRELIMINARY; PRT; 1250 AA.
 AC 088971:
 DT 01-NOV-1998 (TREMBLrel. 08. Created)
 DT 01-NOV-1998 (TREMBLrel. 08. Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16. Last annotation update)
 DE CDO.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA kang J.S., Mulieri P.J., Miller C., Sassoon D.A., Krauss R.S.;

"CDO, a Rbdo-related cell surface protein that mediates myogenic differentiation."

Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.

CC - SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX

CC DOMAIN

DR EMBL: AF090866; AAC43031.1; -.

DR InterPro: IPR001777; -.

DR InterPro: IPR003006; -.

DR InterPro: IPR003598; -.

DR Pfam: PF00041; f03; 3.

DR Pfam: PF00047; i9; 5.

DR SMART: SM00408; Igc2; 1.

DR SEQUENCE 1250 AA; 135362 MW; 7F9AE93D94D8CCF CRC64;

Query Match 2.4%; Score 152; DB 11; Length 1250;

Best Local Similarity 19.5%; Pred. No. 0.061;

Matches 208; Conservative 132; Mismatches 378; Indels 346; Gaps 54;

```

OY 309 PYVMNMPETKARAGOSVSLCCATGKPRDPKRYFMYHNDTLDPGLYK---HESKLVLRK 365
DB 28 PFYIEEPISAVOKLGRPVYLHCSA--KPYTARIIMLHNGKRLDRNTEQIKIHGTLTILS 85
OY 366 LOOHQAGEYFCKAOSDAGAVKSKVAQL-----IVTASDETP---CNPVE 407
DB 86 LNPISLGGCYQAVNNSVGAIVGSPATVSAALGDFDSSFMHVTAEKMTGFIGCR-VPE 144
OY 408 S-----YLIR---LPHDCFONATNSFYDYGRCPVTKACAGQDNGICRDVAONCCGI 457
DB 145 SNPKAEVRKIRIGKMLKH-----STGNYILLPSGNLOVLNVSSEKDKSYKC--AAVN---- 194
OY 458 SKTEEREICSGYTL-----PTKVAKEC--SCORCETRS 490
DB 195 PYTSELAKVEPTGRKLLVSRSSNGPHILPALSOALAVIPHSPTVLECYVSGVPASQVYM 254
OY 491 IYRGVSAADNGEPMRFCHVYMGN-----SRVSMYGYKGFETLHVPODTERLVLEVD 543
DB 255 LKDGDAVAGSMRRLYSHLATASIDPADSGNVCVGNKSGQKVHV-----TYMV 305
OY 544 RLOKRVNTTKVLPFK--KGSVAV-----FHEKMLRKREPTILEMEFN 585
DB 306 NVLEHASISKGLHDKVSLGATVHFTCDVGNPARNRTWFHNAQPI--HPSSRHILEGN 362
OY 586 ILPLGEVGEDMAELEIPSR--SFYRONG-----EPYIGVKASVFLDPRNLS 633
DB 363 VAKITGVWEDSGLYOCVVDNIGIFWQSTGRQIIEODSGMKPIYTAAPANIEWMDGFEVT 422
OY 634 TATAAOTDLNFTNDSGDFRP-LRTYGMFSVDFREVTSEP---LNAKVKVH----- 681
DB 423 LS-----CNATGVPVPVIMHVGRLH-----ITSHPSQVLRKSKPRKSHLFRPDLD 468
OY 682 LIDSTQVKMPEHISTYKLMNLNDTGLMEEGDPKFE-NQRNRKREKRTLVNGLREKR 740
DB 469 LEPVYLIMQAGSS--SLSIQAVTLEHAGKYTCENATKHSQSEALVLT----- 515
OY 741 LFNLLVPESSRCFVKVRAVSERFLPSEOIQ-----GVVSYVNLPRPTGFLSNPRA 792
DB 516 ----VVFETIN-----KAESVTPSEASQNDERPDRGSESSLNLP----- 554
OY 793 WGRFDSVITGPNGACVPARCDOSPDAYS-----AVYVLSIAG----- 830
DB 555 ----VKVHPSGVELPAEKNASVDPAPNILSPQTHMPDNLVLRARCDGMPINAYF 608
OY 831 -----EELQAVESSPK-----FNNMAGVQPIYLNKLIN 858
DB 609 VKYRKILDDSGAVGSMHTVAVPGSENEHLTELEPSLEYVLVAVASAAEGGQPAW--LT 666
OY 859 YRTHDEHEDRVAKT--AFQISMAKPRPNASBESNGPIYAFENLRACEAPPSAHERFYQ 916
DB 667 FRTSKKMASSKNTQASFPVGLPKRPVTAELASNS-----NFGV 705
OY 917 IEGDRDYNTVFPNEDDP-----MSWTEDYLAWMPKP-----MEPRACIYIKYKI----- 960

```

```

DB 706 VLTSSRHSVGEAPDRPTISMASETSYVVTWIPRANGSPITAFKVEYKRMRTSDMLVA 765
OY 961 ----VGP--LEVVNRSRNMGGTTHRRYTKGL--YGIRVRSRRDDQDNVSAACEFQCSG- 1012
DB 766 AEDIPSKLSVEYRSLEPSIKFERVIAINHG---ESFR-----SSASRPVOVAG 814
OY 1013 -----MLYDODRVDTLV---KVIPGSCRRASVNPMLHEYLVLNPLAVNN 1056
DB 815 PNFRSNRPITGPHIATEAVSDQIMLKMTYPS-----SNNTPIQGFYIYRPTDSN 869
OY 1057 DTSEYTMALPDLPLGNHYGIYVTDODPTAKEIALGRCFDGTSQSSRKMSNVGVALT 1116
DB 870 D-SDYKR----DVVEGSKOMHTIGHIQPETSVDIKM-QCFN--EGGESEF--SNVICET 919
OY 1117 FNCVERQVGRS-AFOYLDSTPAQSPACTVQGRVSRROQARS 1159
DB 920 --VKRVPGASDYPVKELSTPSSSGMAGNVG--PATSPARS 958

RESULT 11
O76281 PRELIMINARY; PRT: 6658 AA.
AC 076281; Q24343;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE BEN1 PROTEIN (FRAGMENT).
GN BT OR CG1479.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98300339; PubMed=9636710;
RA Daley J., Southgate R., Ayme-Southgate A.;
RT "Structure of the drosophila projectin protein: isoforms and
RT implication for projectin filament assembly.";
RL J. Mol. Biol. 279:201-210(1998).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=91376068; PubMed=1910171;
RA Ayme-Southgate A., Vigoreaux J., Benjan G., Pardue M.L.;
RT "Drosophila has a twitchin/titin-related gene that appears to encode
RT projectin.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:7973-7977(1991).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=95146546; PubMed=7844153;
RA Ayme-Southgate A., Southgate R., Saide J., Benjan G.M., Pardue M.L.;
RT "Both synchronous and asynchronous muscle isoforms of projectin (the
RT Drosophila bent locus product) contain functional kinase domains.";
RL J. Cell Biol. 128:393-403(1995).
CC -1 SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1 SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN
DR EMBL: AF047475; AAC27550.1; -.
DR HSSP: P00518; 1PK.
DR FLYBase; FBgn0005666; bt.
DR InterPro: IPR000255; -.
DR InterPro: IPR000577; -.
DR InterPro: IPR000719; -.
DR InterPro: IPR001245; -.
DR InterPro: IPR001777; -.
DR InterPro: IPR002290; -.
DR InterPro: IPR003006; -.
DR InterPro: IPR003600; -.
DR Pfam: PF00041; f03; 37.
DR Pfam: PF00047; i9; 16.
DR Pfam: PF00069; PKinase, 1.
DR PRINTS; PR00014; FNTYPEITI.
DR PRINTS; PR00109; TYRKINASE.

```

DR PROSITE: PS00933; EGGY KINASES.1; UNKNOWN.1.
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; UNKNOWN.1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR SMART: SM00410; IG_Like; 1.
 DR ATP-binding; Repeat; Serine/threonine-protein kinase; Transferase;
 KW Tyrosine-protein kinase.
 FT NON_TER
 SQ SEQUENCE 6658 AA; 743460 MW; 0B7ZEB0993F6062E CRC64;

Query Match 2.4%; Score 152; DB 5; Length 6658;
 Best Local Similarity 19.6%; Pred. No. 0.08;
 Matches 162; Conservative 100; Mismatches 259; Indels 306; Gaps 40;

QY 234 GGAPASGAII-----YLLKTPKLLTQTD-----SDGRFRIPGLCPDGK---- 272
 DB 1742 GGASITGYIVERKDDPNTGKWKQKALETSFDPCKARVNDLIAGNKQYFRIMAVKAKGSKPS 1801
 QY 273 --SLIKITKVKFAPIVLTPMK---TSIKRATIKAEFVRAETRYMWNPETKARRAGQSVS 327
 DB 1802 EPSDQMTAKDRAP-----PKIDRTNIDKITSK-----AGQHTR 1835
 QY 328 LCCATGKPRDPKYFWYHNDTL--DPSLY-----KHESKIVLRKIQOHQAGEYPCXKAS 380
 DB 1836 FDIKVSGER-PATKYWLHNKARLENDSDSNYNDMESYRKILVPTISKRHSKITLKAEN 1894
 QY 381 DAGAVKSKVAQILIVASDETPCNPPVPESTILRLPHDFCFQMATNSFYVDGR--CPVKTC 438
 DB 1895 ESGRDEASFEVIVLD-----KCPGPPGPLERVY-----DVHKRGCKLKNWA 1934
 QY 439 GOODNGICRDAVQNCGCSIKTEEREIOCSGTLPTKYAKESQ--RCTETSTYRGRYS 497
 DB 1935 PLDDGGLIDHYI-----LEKMDVE-SGHWLPSGRFKESFALNNLEPSHEXKFRVL 1985
 QY 498 AAD--NCEPMKFGHYVGNRSVMTGYKGTFTLHVPODTERIVLTFYDRLQKFNVTYKV 554
 DB 1986 AVNTEGESEPLTGEQSVLAKNPFDEPKRGY-----PE-----AVDMQKDHVDLYWR 2032
 QY 555 LPPNKKGSVAVHEIKMLRKKEPT--LEAMETNIIPAGE-----VVGEDPMALLET- 603
 DB 2033 PPIINDGSGPIIGYV--YKEREKGTDMKIKGTETITPCLEGECATVPTLNENCEYEFYVK 2090
 QY 604 -----PSRSFYRQNGEPIYGV--KASYTFIDP--RNISTATAQIDLNF--INDGCD 650
 DB 2091 AINAGREPSDA-----SKPIITKPKLAPLIDPYNINRTYNEFKSGEPIFLDINISGE 2145
 QY 651 TEPILRT-----YGMFSYVDR 665
 DB 2146 PAPDVTWNNKNSVQTSFSHIENLPYNTKYINNPERKDTGLKISAHNFYGGDQVDFQ 2205
 QY 666 DEVISEPLNAG---KKVHLDSIQVKNP-----EHISTVKLMSLNDPTGLMEEDGF 714
 DB 2206 INITTKGKPGGPLEVSEVHKDKLKKWKPKDGGEFVESYLEVEKPPDPTGIMLPVG-- 2263
 QY 715 KFEQQRNKRREDTFLVGNLEIRERRLEFNLDVPSRCFVAVRYRSRFLPSQIQGV 774
 DB 2264 -----RSDGPE-----YNDGLVPGHDYK--FR 2284
 QY 775 ISVINLEPRTGFLSNPRAMGRFDSVITGPNGACVPAFCDDSPDAYSAVYLASILEQ 834
 DB 2285 VKAVNKE-----GESEPLET-----LGSIIADPF 2309
 QY 835 AVESSEPFENRAIVPOPYLKNLKYRRTDHEDPVKKTAFOISNAKPPNSAEBSNGPI- 893
 DB 2310 SVLPKP-----GVPP-----TDW-----TANKVELAMPBP--ASDGGSPILQ 2344
 QY 894 -YAFENT-----RACEAPPSAAHFRFYQIEGDRYDNTVPFNE 931
 DB 2345 GYIYEVVDKISPLMEKALETNSPTPTATVQGLLEGNEYQFRVALNK 2391

RESULT 12
 ID 075851 PRELIMINARY; PRT: 4123 AA.
 AC 075851;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE WUGSC:H.DJ0751H13.1 PROTEIN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OC NCBT_Taxid=9606;
 RN [1]
 RA SEQUENCE FROM N.A.
 RP Leonard S., Graves T., Strommatt C.;
 RT "The sequence of Homo sapiens PAC clone DJ0751H13";
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RA SEQUENCE FROM N.A.
 RP Waterston R.H.;
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RA SEQUENCE FROM N.A.
 RP Waterston R.;
 RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
 CC -1 SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
 (LDLRA) DOMAIN.
 DR EMBL: AC004877; AAC36301.1; -
 DR HSSP: P01130; IAUJ.
 DR InterPro: IPR000421; -
 DR InterPro: IPR000561; -
 DR InterPro: IPR000753; -
 DR InterPro: IPR000884; -
 DR InterPro: IPR000923; -
 DR InterPro: IPR001007; -
 DR InterPro: IPR001064; -
 DR InterPro: IPR001092; -
 DR InterPro: IPR001846; -
 DR InterPro: IPR002172; -
 DR InterPro: IPR002223; -
 DR InterPro: IPR002919; -
 DR Pfam: PF000057; 1dl_recept_a; 11.
 DR Pfam: PF000090; tsp_1; 14.
 DR Pfam: PF00094; vwd; 5.
 DR Pfam: PF00754; F5_P8_type_C; 1.
 DR Pfam: PF01826; TIL; 9.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PROSITE: PS00280; BPTL_KUNITZ; UNKNOWN.1.
 DR PROSITE: PS00196; COPPER_BLUF; UNKNOWN.1.
 DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN.1.
 DR PROSITE: PS00225; EGF_1; UNKNOWN.1.
 DR PROSITE: PS01209; LDLRA_1; 9.
 DR PROSITE: PS50068; LDLRA_2; 20.
 DR SMART: SM00011; VWC_def; 1.
 KW Glycoprotein.
 FT NON_TER
 SQ SEQUENCE 4123 AA; 434985 MW; 7AAB6FE8DCE012FB CRC64;

Query Match 2.4%; Score 150; DB 4; Length 4123;
 Best Local Similarity 21.3%; Pred. No. 0.05;
 Matches 117; Conservative 40; Mismatches 192; Indels 200; Gaps 30;

QY 81 YGDRVCARPLRLAETDWTDPAGSTGOVHSGSPREGFW-----C-----LNRE. 123
 DB 3028 FGGAEGCGP-TMEAFPSILRC--PGPYPGMCPDCKMDLDCAGPASCALSLARGTNQT 3084
 QY 124 QRPQNCNSNTVRLCP---PGSLRDPTEI-----WSPWSEWSKCSA 163
 DB 3085 CHPCCHCPSGMLTLVSRGHGPGAGASQPPVALPGAIGTSVAGAGMGPMGPWSHCGR 3144
 QY 164 ACGGTGYOTRTRIC-----LAEWVSLCSEASEGQCHMG-----QDCTA----- 202

```

Db 3145 SCG-GLSRRTACDPPPOGLDYCEGPRAQEVQALPCPVTNCTAIEGAEYSPGCP 3203
OY 203 -----CD-----LTCPMGOVNMADCA-----CMQODPMLHANVSLPGAPAS 239
Db 3204 CPRSCDDLVHCWRCQPCYCPCGGVLSNGAICVQPGHCCLD-LLTQORHHPGAR--- 3259
OY 240 GAATVLTPTKPLTQTDSDGRFRIPLGL-CPD---GKSLTKTKVFAFIVLTPMTSLK 295
Db 3260 -----LARPDCNHCCTGLEGRLNCTDLPCPDGGGSL-----HCCGPPCRPS--- 3302
OY 296 AATTAEEFYRAETPYVMNPETKARRAGOSVSLCCAT-GKRPDKYFMYHNDTLDPSL 354
Db 3303 -----CQDLSFGSVSCPG-----SVGCPCTGCGPLGQ-----LSQDGLCVPPA 3340
OY 355 YKHSKIVLRKIQOHQAGYFPCAKSDAGAVSKYAQLVTASDERPCNPYESYLIRLP 414
Db 3341 H-----CRQYQPGAMAPSPVPSCTVAG-ILQCEVP----- 3371
OY 415 HDGFONATNSFYDYVGRCPVKTCAQDQNGICRDVAVNCCGISKTEEREICSGYTLPT 474
Db 3372 -DCPPPGVWSSMGPEWDCSVSGSGEQLRSRRC--ARPPCP-----PA 3412
OY 475 KVAKSCSCQCTETRSIVGRV---SAADNGEPMRPGHVMGNSRYSMTGYGTFT---- 527
Db 3413 ROSRSTQVCHREA-GCPAGRLYRCQPGEGCPFSCAHV-----TQVGCSEGE 3462
OY 528 --LHVPQDT 534
Db 3463 ECHCPDEPT 3471

RESULT 13
ID 022631 PRELIMINARY: PRT: 788 AA.
AC 022631.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE T21B6.3 PROTEIN.
GN T21B6.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Cottage A.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ritken L., Koopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spoot J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z68011; CAA92014.1; -.
DR InterPro: IPR000884; -.
DR InterPro: IPR002965; -.
DR InterPro: IPR003609; -.
DR Pfam: PF00090; tsp_1; 6.
DR PRINTS: PR01217; PRICHEXTENS.
DR SMART: SM00473; PAN_AP; 1.
SQ SEQUENCE 788 AA; 88727 MW; BB0DD9F1D29BD961 CRC64;

```

```

Query Match 2.3%; Score 147.5; DB 5; Length 788;
Best Local Similarity 27.2%; Pred. No. 0.0069;
Matches 55; Conservative 15; Mismatches 65; Indels 67; Gaps 11;

OY 46 PADLTPSEWMTLTFNIDPVGKGYERLDAIRF-----YYDRVCARPLRL 93
Db 501 PCQTMSEWMT-----CSASCSGQRER--TRFCLGTNRCEKDYEEQCSAGCPPEM 554
OY 94 ARTTWTGAGST-GQVHSGPREGEWCL-----NBRQ--PGQNSNTVRF 138
Db 555 SQMEDWGQSVTCGQGVAVRQRT---CLGVEFDHLQGPRTGQRACDGPSCS----- 604
OY 139 CPPLSLRRTERTIRWSPWMSKCAAGOTGYOTRTRICLAEVSLCSASEGQCHMG 198
Db 605 -----LWSPQWSEWSTCSASCG--SGMKRRQVRC--QFGTDCGPNESQPCYGP 649
OY 199 DC-----YACDLTCPMQ 211
Db 650 PCAEWTEWCEWMSGCSKCGPGQ 671

```

```

RESULT 14
ID 017591 PRELIMINARY: PRT: 1444 AA.
AC 017591.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE C02B4.1 PROTEIN.
GN C02B4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurtry A.;
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z50004; CAA90283.1; -.
DR EMBL; Z50006; CAA90293.1; JOINED.
DR EMBL; Z50006; CAA90302.1; -.
DR EMBL; Z50004; CAA90302.1; JOINED.
DR InterPro: IPR00130; -.
DR InterPro: IPR000884; -.
DR InterPro: IPR001590; -.
DR Pfam: PF00090; tsp_1; 12.
DR PROSITE: PS0215; ADAM_MERO; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SMART: SM00209; TSP1; 1.
SQ SEQUENCE 1444 AA; 16067 MW; 32D065E59FABBF75 CRC64;

```

```

Query Match 2.3%; Score 147; DB 5; Length 1444;
Best Local Similarity 28.9%; Pred. No. 0.019;
Matches 44; Conservative 22; Mismatches 64; Indels 22; Gaps 8;

OY 86 CARPLRLAERT-----TMTTP-AGSTGOVHSGPREGEWCLNREORPQONCSNYTVR-F 137
Db 1297 CPDPACHLRDGGWSWWSMTPCASCGVQGRDSS---CSSPEKGGQSGSLAHQTS 1353
OY 138 LCPPGSLRDTERTIRWSPWMSKCAAGCOTGVQTRTRICLAEVSLCSASEGQCHMG 197
Db 1354 LCDLPACDHESDGEWASANNWMSGCMGNC-IGTRTVACVSPV-----SDGGQPCFG 1406
OY 198 ODCTACDLT-CPMGVNVADCDACMCDFMLHG 228
Db 1407 R---SSEITECROSPSTALCSSFTITSSHLADG 1435

RESULT 15
061541

```

